





NEW RIVER STATE SCHOOL  
LIBRARY

Division 11.2

Number 11



MEASUREMENT AND  
ADJUSTMENT SERIES

EDITED BY LEWIS M. TERMAN

STATISTICAL METHOD  
IN EDUCATIONAL  
MEASUREMENT

BY ARTHUR S. OTIS, PH.D.

Formerly Development Specialist with  
Advisory Board of General Staff, U. S. War Department  
Author of Otis Group Intelligence Scale  
and other tests



Yonkers-on-Hudson, New York  
WORLD BOOK COMPANY

2126 Prairie Avenue, Chicago

1925

LB  
3063  
06

# WORLD BOOK COMPANY

THE HOUSE OF APPLIED KNOWLEDGE

Established 1905 by Caspar W. Hodgson

YONKERS-ON-HUDSON, NEW YORK

2126 PRAIRIE AVENUE, CHICAGO

---

Scientific method in education involves the careful measurement of each child's ability to learn and of the amount that he has learned. It also involves adjustment of organization, subject matter, and methods of instruction to the varying needs and abilities of pupils. This book is one of a series that sets forth the value, technique, and applications of educational measurement and adjustment. It describes in an elementary way those simple and useful methods and devices that are commonly employed in the interpretation of the data of mental and educational measurements

MAS : OSMEM-1

---

Copyright 1925 by World Book Company

Copyright in Great Britain

*All rights reserved*

---

PRINTED IN U.S.A.

## PREFACE

THIS book deals with those simple statistical methods that are needed in the interpretation of test results. It is designed for classroom teachers, administrators, students, and researchers.

The purpose of the book is to present the subject in such a way that it can be understood by those who know nothing whatever about statistical method. The reader is given an insight into the subject so that he will see the reasons for the various kinds of statistical procedure and will understand the meaning and significance of the results.

The attempt has been made to cover all the topics that a school teacher, administrator, or researcher is likely to have need for, but only those, the reader being referred to more technical books for explanation of special procedures that are seldom used. Particular attention is given to the explanation of the meaning and significance of correlation, and the simple and useful applications of partial and multiple correlation are explained.

The reader somewhat familiar with statistical method and terminology will find frequent footnotes which qualify, in the interest of exactness, the more general statements of the text. The lay reader need not trouble to study such notes, for they are not essential to a working knowledge of the method.

Several new charts for practical use are introduced, including a correlation chart and a percentile graph. These make easy certain important processes and calculations once known only to specialists in statistical method. The past few years have shown a very definite trend toward the development of such practical devices, and their more extensive use in future is certain; for this reason considerable space is devoted to the explanation of them.

Indeed, throughout the book the practical application of

methods has been kept in mind, and only such discussion of theory is given as is thought necessary to make the use of the methods and devices intelligent rather than rule-of-thumb. For example, the reader is shown the meaning of correlation by very simple illustrations with all refinements omitted, whereas the application of the complete formula is reduced by "job analysis" to a mere succession of simple arithmetic steps which may be performed with no thought of the formula. Attention may be centered on accuracy of the computation, which is reduced to a minimum by various tables.

It is believed that any person interested in scientific method in education may obtain from this book a good working knowledge and understanding of statistical method, no matter how new the subject may be to him.

#### ACKNOWLEDGMENTS

The writer is indebted to Lewis M. Terman for a very careful and helpful editing of the book, and to the following persons for reading the manuscript and offering valuable suggestions for its improvement: John R. Clark, Truman L. Kelley, William A. McCall, J. Ralph McGaughy, Rudolf Pintner, Godfrey H. Thompson, and Ben D. Wood.

John R. Clark also read the proof and has given especially helpful criticism.

Mrs. Otis has spent many patient hours with proof reading, computations, and copying of drawings.

Thanks are due the following persons who have contributed material for illustrations, exercises, etc.: E. J. Ashbaugh, George T. Avery, H. R. Bonner, Leo J. Brueckner, B. R. Buckingham, B. P. Heubner, A. M. Jordan, J. Cayce Morrison, H. H. Murphy, W. M. Proctor, Percival M. Symonds, Jacob Theobald, Elmer H. Webber, and Sara E. Weisman.

ARTHUR S. OTIS

# CONTENTS

	PAGE
EDITOR'S INTRODUCTION . . . . .	xi
 CHAPTER	
I. INTRODUCTION . . . . .	1
Is there anything interesting about statistics? — The purpose of this book — Is there a science of mental measurement? — What kind of facts does statistical method reveal?	
II. THE CENTRAL TENDENCY OF A DISTRIBUTION . . . . .	6
Comparison of groups — The mean — The median — Making a distribution — Range of score — Overlapping — Finding the median from a distribution — Special cases — Finding the mean score from a distribution — The mode — Summary	
III. REPRESENTING THE WHOLE OF A DISTRIBUTION . . . . .	20
Comparison of individuals within a group — Rank order — Comparison of pupils of different groups — The percentile rank — An approximate method of finding a percentile rank — Interpreting the rule — The bar graph — The histogram — The step graph — Finding the median from a step graph — Quartile scores — The line of diagonals	
IV. GROUPED DISTRIBUTIONS . . . . .	36
Need for a new method of dealing with scores — Finding the mean of a grouped distribution — Histograms — The graph of a grouped distribution — Finding the median of a grouped distribution — Finding the median graphically — Finding quartile scores graphically — Calculating the median of a large group — A point of theoretical interest — Some theoretical considerations — Questions	
V. THE PERCENTILE GRAPH . . . . .	53
The percentile graph — The general utility of the percentile graph — How to draw a line of diagonals on a percentile graph — Cumulative errors — A table of per cents — Plotting points by means of the scale chart — Drawing the line of diagonals — The Universal Percentile Graph — Drawing two or more lines on one graph — The smoothing effect of grouping — The general shape of a graph — Finding a percentage graphically — Questions	
VI. THE LAW OF NORMAL DISTRIBUTION . . . . .	68
The law of probability in the flipping of coins — The chance element in scores — The normal surface of distribution — The normal curve of distribution — Skewed distributions	

CHAPTER	PAGE
VII. PERCENTILE CURVES . . . . .	77
The nature of a sample — Effect of smoothing a line of diagonals — Percentile curves — The difficulty of smoothing a histogram — Advantages of the percentile curve	
VIII. VARIABILITY OF SCORES . . . . .	85
Need for measuring of variability — The total range — The interquartile range — Overlapping of classes — Other measures of variability — Median deviation — Probable error — Average deviation — Standard deviation — Variability of normal distributions — Questions	
IX. PERCENTILE RANK IN A NORMAL DISTRIBUTION . . . . .	95
The correspondence between scores and percentile ranks — Percentile rank in terms of variability — Testing the normality of a distribution — An accurate table of percentile ranks	
X. THE CORRESPONDENCE BETWEEN TESTS . . . . .	101
The problem — A method of finding the correspondence between scores in two tests — Making a table of correspondence — Use of percentile graph in finding correspondence — The line of relation — Alternative methods	
XI. AVERAGING MEASUREMENTS OF A PUPIL . . . . .	118
How to average scores in two tests — Disadvantages of averaging percentile ranks — The method of transmuting scores — The need for a universal standard scale — <i>T</i> scores — Disadvantages of the <i>T</i> -score method — How to average scores and teacher's marks	
XII. THE GROWTH OF MENTAL ABILITY . . . . .	132
A growth chart — Comparison of growth curves — Growth in mental ability — Validity of mental-ability test scores — Units of measurement of mental ability — A curve of growth in mental ability — Interpreting the curve of growth — Individual differences — Bright, normal, and dull children — Mental age — Norms — The limit of true mental ages — Binet mental ages — Fictitious mental ages — Questions	
XIII. THE MEASUREMENT OF BRIGHTNESS . . . . .	148
The intelligence quotient — The invalidity of the old IQ — The remedy — A popular misconception — The Coefficient of Brightness — Intelligence quotients from group tests — The Index of Brightness — A new method of finding an intelligence quotient — Percentile rank — The Interpretation Chart — Finding an intelligence quotient by means of the chart	

CHAPTER	PAGE
XIV. NORMS . . . . .	160
Need for representative sampling — Age norms — Grade norms — How to obtain grade norms from age norms — Grade status — Finding grade status from age norms — Finding age norms from grade norms — Interpolation — Subject ages — Fictitious subject ages — Educational quotient — The accomplishment ratio — Summary	
XV. THE MEANING OF CORRELATION . . . . .	175
The comparison of relative standing — The meaning of correlation — The natural method — Perfect correlation — Coefficient of correlation — Significance of the coefficient of correlation — A coefficient of correlation in relation to causation — Questions	
XVI. THE CALCULATION OF A COEFFICIENT OF CORRELATION . . . . .	186
The product-moment method — The difference method — The difficulty with fractions — The use of an assumed mean — The Otis Correlation Chart — How to use the Correlation Chart — Negative correlation — Rank methods of computing correlation — The second rank method — The Spearman Footrule — Interpolation — Correlation by unlike signs — Other methods of measuring correlation — Questions	
XVII. CORRELATION AND PROGNOSIS . . . . .	217
The criterion — The coefficient of alienation — Further interpretation of a coefficient of correlation — Interpretation of a coefficient of correlation in terms of displacement — Errors of measurement — The effect of errors of measurement upon correlation — Reliability — Correlation for attenuation — Questions	
XVIII. PARTIAL CORRELATION . . . . .	230
A prevalent misconception regarding correlation — The need for a more direct measure of correlation — How to find a coefficient of partial correlation — Nature of partial correlation — Effect of heterogeneity on correlation — How to correct a coefficient for a change in heterogeneity — Question	
XIX. MULTIPLE CORRELATION . . . . .	238
Measuring prognostic value — The meaning of multiple correlation — The formula for multiple correlation — Finding the best weighting — Formula for weighting — Symbolization — Relation of a multiple correlation to the total correlation — Regression equations — Calculations with four or more variables	

CHAPTER	PAGE
XX. RELIABILITY . . . . .	247
<p>The fallibility of test scores — Causes of variability of the scores of a test — Reliability — Measures of reliability — Probable error of a test — How to find the probable error of a test — The formula for finding the probable error of a score from the differences between pairs of scores of single individuals — Need for other measures of reliability — Probable error in terms of variability of scores — Reliability coefficient of correlation — Relation between reliability coefficient and probable error — Formula for finding the probable error of a score from the reliability coefficient — Interpretation of a reliability coefficient — Reliability and heterogeneity — Errors of measurement — Validity — The probable error of a coefficient of correlation — The formula for the probable error of a coefficient — How to find the probable error of a coefficient of correlation obtained by the Otis Correlation Chart — Significance of the probable error of a coefficient — The probable error of a difference — Probable errors of a mean and of a standard deviation — Probable error of coefficients of partial and multiple correlation — Reliability of averages — Practice effect — Question</p>	
XXI. GRADING AND CLASSIFYING . . . . .	267
<p>The problem — Regrading on the basis of mental ability — Need for an achievement test — Regrading by counting papers — Regrading by a percentile curve — A less drastic method of regading — Classification within grades — Bright, normal, and dull sections — The need for varied courses — The three-track plan — The Interpretation Chart — Convenience of the Interpretation Chart — Classification for the three-track plan — Grading within bright, normal, and dull groups — No rigid classification — Taking account of teachers' marks in grading and classifying — How to take account of teachers' marks — Rating scale — A "five-point" scale</p>	
APPENDIX :	
I. Abbreviations and Symbols . . . . .	289
II. Statistical Tables . . . . .	291
III. Supplementary Practice Material . . . . .	306
IV. Answers to Exercises . . . . .	317
V. Bibliography . . . . .	321
INDEX . . . . .	329
AGE CALCULATOR . . . . .	333
IQ SLIDE RULE , , , , , . . . . .	337

## EDITOR'S INTRODUCTION

IN recent years extraordinary changes have taken place in the professional training of teachers, as a result of the development of scientific methods for the measurement of intelligence and educational achievement. Textbooks devoted to arm-chair discussions of educational principles, child study, and general methods are rapidly giving way to texts which deal with factual data resulting from the measurement of native abilities and of the effects produced upon such abilities by particular educational influences. This does not mean that the larger issues of educational theory and practice are being neglected, but only that we now recognize the futility of any attempt to dispose of such issues except in the light of investigational findings. At every point the consideration of educational theory must wait upon the discovery of facts regarding the nature of the raw material with which teachers work and regarding the changes which various kinds of teaching methods are capable of making in this raw material.

Present-day methods of educational investigation necessitate the constant use of statistical methods in the treatment of data. Familiarity with statistical procedures has for some time been considered necessary for the psychologist and the educational research worker, but we are rapidly coming to recognize that it is also an indispensable part of every teacher's equipment. Teachers cannot be expected to make intelligent use of test methods as long as the significance of test results is hidden from them in a maze of meaningless figures. But while every one will admit the *desirability* of acquainting the average teacher with the mysteries of statistical procedures, some may be inclined to doubt its possibility. Certainly it is impossible by the use of any of the textbooks which have hitherto been available. There are several statistical treatises which are reasonably suitable

for use with advanced graduate students of psychology and education, but in the editor's opinion this is the first text that is at all satisfactory for use in teachers' colleges or in teachers' reading circles. It is at the same time admirably adapted for use as an introductory text in colleges and universities.

It is generally recognized that the course in statistical methods presents, from the pedagogical point of view, exceptional difficulties. These are traceable in part to the inadequate mathematical preparation of the average student of education, but in part, also, to faults of exposition on the part of the textbook and the teacher. Principles which could be made concrete are presented abstractly; sequence is disregarded; insufficient drill is provided in fundamental everyday procedures. The average student who has taken such a course may have acquired a little "knowledge about" statistical methods, but he can make little or no use of them. The present textbook by Dr. Otis will do much to remedy this situation. Its characteristic features lie in the fact that it takes nothing for granted, that it relies upon the simplest and most straightforward explanations, that it gives concrete meaning to abstract terms and principles, that it teaches statistical procedure in connection with data of universal interest to teachers, and that it provides at every stage the practice necessary for making the procedures habitual. Students and teachers who have studied this text will be able to apply accepted statistical methods with all the ordinary kinds of measurement data, and will be vastly more able to read and understand the current literature of education and psychology.

Dr. Otis is widely known among American teachers as the author of several of the most useful tests of intelligence and of educational achievement. Among a smaller group he is known for his important contributions to statistical theory.

The book here presented rests, therefore, upon a solid foundation of mathematical knowledge and of practical experience in the use of measurement methods. From the pedagogical point of view it is without a rival in its field. It will add to the popularity of tests because it makes their results meaningful to the non-mathematical.

LEWIS M. TERMAN



# STATISTICAL METHOD IN EDUCATIONAL MEASUREMENT

## CHAPTER ONE

### INTRODUCTION

Is there anything interesting about statistics? Most people would probably say no. They mean that most statistics are for them uninteresting. This is quite natural. But when a mother keeps a record of the growth of her baby from week to week, or when a teacher who suspects that her pupils are not as bright as those of another teacher undertakes to find out by testing them with a mental-ability test whether this is true, these data become for that mother or teacher very interesting.

**The purpose of this book.** In reality, this book does not deal with statistics as such. It deals with the method of interpreting statistics — one's own interesting data, collected painstakingly for an important purpose. When the teacher has taken anywhere from one hour to ten hours in giving and scoring a standard test and tabulating the scores, she wishes to get from these data all the significance and enlightenment that can be found in them. This book is written for the special purpose of enabling a teacher who knows nothing whatever about statistical methods to work over her scores and any other data in the most convenient way, so that those facts that are inherent in the data and which are earnestly sought will stand out vividly and in their true perspective.

After all, those things are likely to be most interesting about which we know most, and the writer firmly believes that the more the teacher or principal learns about the methods that may be invoked to make a seemingly incoherent mass of scores or other measurements yield fruit in those truths that are so vital to the progress of education, the more

## 2 *Statistical Method in Educational Measurement*

interest that teacher or principal will find in the study of the method itself.

Is there a science of mental measurement? When one speaks of science we generally think of some organized system of relatively exact knowledge, such as mathematics, physics, or chemistry. Mental measurement, on the other hand, is so inexact that we may be inclined to feel that it should not be called a science. However, the method of handling the data of mental measurement statistically in order to determine how inexact they are and what degree of reliance may be put upon them constitutes one of the most exact of sciences. We may have used a very inadequate measure of the ability of a pupil in written composition, for example, but by the collection of a large number of such measurements we are able by statistical methods to determine the most probable true measure of a given pupil's ability and to know just how much reliance may be placed upon the results.

What kind of facts does statistical method reveal? The statistical method described in this book is equally applicable to the measurement of "intelligence," of the products of education and training, of character traits — indeed, of any kind of anthropometric measurements. The emphasis in this book, however, is laid on the application of the method to so-called mental and educational measurements.

Let us suppose that a school principal has administered a mental-ability test and an achievement test to the pupils of Grades 4 to 9 of his school. What may the scores reveal?

(1) The principal may determine, within a known degree of accuracy, the relative amount of the mental development of each pupil tested and the relative amount of knowledge he has attained during his school career.

(2) He may determine, within a known degree of accuracy, the average level of mental development of the pupils of any

grade in comparison with that of the pupils of any other grade, and determine whether the steps in average development from one grade to another are approximately uniform. He may do the same for the steps in average amount of knowledge possessed by the pupils of the several grades.

(3) He may determine how widely different in amount, or, as we say, how *variable*, the abilities of the pupils of a single class are, and may compare the variability of one class with that of another. Stating this in another way, we may say that the principal may determine the relative *homogeneity* of his various classes in mental ability and in knowledge (achievement).

(4) He may determine the degree of *overlapping* of grades and classes in mental ability and achievement. For example, he may find that 25 per cent of the pupils of the fourth grade exceed the average of the fifth grade in achievement, and that 5 per cent of the fourth grade exceed the average of the sixth grade.

(5) If other schools in his city have used the same tests, he may determine how his school stands in comparison with those schools with respect to the average mental development of the pupils and their average achievement.

(6) By means of the *grade norms* given in the manuals of directions, he may determine how his school stands in comparison with schools in other cities with respect to the average mental development of the pupils and their average achievement. (The norm for the fifth grade, for example, is the average score of fifth-grade pupils all over the country.)

(7) He may compare the mental ability of any individual pupil with that of the other pupils of the same grade or class in a number of ways. He may say, for example, that George has a degree of mental ability that places him in the upper quarter of fifth-grade pupils, or he may say that George exceeds in mental ability 85 per cent of the pupils of the fifth grade. There is a simple way of determining this percentage

#### 4 *Statistical Method in Educational Measurement*

by means of a device called the Percentile Graph, which is described in Chapter V. A pupil's relative position in achievement may be similarly expressed.

(8) A pupil's ability may be expressed also in terms of the age for which such ability is normal or average. Thus, the principal may find that George has a mental ability just normal for the age of 12 years, 8 months. (If this is true, George is said to have a mental age of 12 years, 8 months.)

(9) A pupil's achievement may be expressed, if desired, in terms of grade status; that is, if a pupil has made a score in the achievement test which is just average for the fifth grade at the end of the third month of the school year, we may say the pupil has a grade status of 5.3.

(10) The principal may also compare the achievement of each pupil with his mental ability and thus determine which pupils are doing school work that is on a par with their mental ability. The ratio between achievement and mental ability may be expressed mathematically as the *Accomplishment Ratio*, as is explained on page 172.

(11) The principal may find the degree of correspondence between achievement and mental ability by the method of *correlation*. This may be done by means of a Correlation Chart, which is described herein (pages 192 to 201). The chart is designed for use by persons wholly unfamiliar with the method of correlation. By the use of this method the principal may find also the degree of correspondence between achievement as measured by the test and achievement as judged by the teacher, or between any two abilities, either measured or estimated.

(12) If one form of the test is given at the beginning of the year and an alternative form at the end, the principal may determine the relative progress made in achievement by each pupil and by the various classes. By means of measures of progress, it is possible to compare the merits of various

teaching methods and devices. In scientific procedure of this sort rests the future development of educational method.

In enumerating the various items of information that a school principal may discover from the results of the administration of mental and educational tests, we have not mentioned the *application* of these, chief of which, of course, is the classification of pupils into homogeneous groups in order that instruction may be better adapted to the needs and abilities of the pupils. Also, it is not to be inferred, of course, that such discoveries may be made only by the principal. Any teacher or supervisor may make the same investigation as far as her own pupils are concerned.

The determination of any one of the facts mentioned above regarding one's pupils, however, requires the understanding of statistical method, at least to some small degree. It is the aim of this book to set forth in an elementary way those principles and procedures which will enable any teacher, supervisor, principal, superintendent, or other researcher to obtain regarding his pupils any of the kinds of information mentioned above, and many others, without his having made any previous study of statistical method.

It may be advantageous to read each chapter through as a whole without attempting to master it as you go along. In Chapter III, for example, several methods of finding a percentile rank are discussed, first according to a common conception, next a precise but rather impractical method, and finally an approximate method that is simple and practical. You should become familiar with the third method, and for that purpose an exercise is provided giving practice in finding a dozen percentile ranks. It is not necessary, however, to spend time studying the first two methods unless you are especially interested in these. In general those discussions not followed by exercises are of a type leading to more practical methods for which exercises are given.

## CHAPTER TWO

### THE CENTRAL TENDENCY OF A DISTRIBUTION

**Comparison of groups.** The pupils of a certain school in New York were recently tested with an arithmetic reasoning test. The scores of the pupils of Grades 5A and 5B are given in Table 1.

Is it possible to tell from inspection of this table whether there is a tendency for the scores of one class to run higher "on the average" than the scores of the other? Make a guess from inspection of the table as to which grade has made the better scores "on the average." Guess about how much better the scores of one grade are than those of the other.

TABLE 1  
SCORES OF PUPILS IN GRADES 5A AND 5B IN THE ARITHMETIC  
REASONING TEST

GRADE	NO. OF PUPILS	SCORES
5A	36	11 8 7 8 7 7 5 4 8 10 11 9 10 12 9 6 15 7 8 5 12 14 11 5 5 9 10 12 4 8 7 10 8 8 10 9
5B	47	7 12 8 10 11 6 11 11 11 10 8 5 11 8 8 6 13 7 9 13 11 9 12 9 6 8 8 8 6 12 6 14 10 16 12 13 10 11 12 10 14 10 10 9 8 13 11

**The mean.** A common measure of the achievement of a class as a whole is the "average" score of the class. This is found by adding all the scores of the class and dividing the sum by the number of pupils in the class. Thus, the sum of the scores of the pupils in Grade 5A is 308. Dividing this sum by 36, the number of pupils, we get  $8\frac{20}{9}$ , which we can call 9 if we wish only the nearest whole number. Carried to two decimal places, the "average" for the class is 8.56

points. The "average" score of Grade 5B is 9.85, or roughly 10. This shows Grade 5B to be ahead of Grade 5A in arithmetic reasoning. (In this school 5B was the high fifth.)

Statisticians generally speak of the "average" score as the *mean* of the scores, or *mean score*. The term *mean*, therefore, when not qualified, always refers to what is commonly known as the *average*; that is, the mean of a series of scores is the sum of the scores divided by the number of scores.<sup>1</sup>

**Formula for the mean.** Another way to state the definition is as follows:

$$\text{Mean of a series of scores} = \frac{\text{sum of scores}}{\text{number of scores in series}}.$$

Let us boil this down still farther.

Let  $M$  stand for the mean.

Let the Greek letter  $\Sigma$ <sup>2</sup> stand for "the sum of."

Then  $\Sigma s$  will stand for "the sum of the scores."

Let  $N$  stand for the number of scores in the series.

The mean of a series of scores is then defined by the following formula:

$$M = \frac{\Sigma s}{N}. \quad \begin{array}{l} \text{(Formula 1, for} \\ \text{the mean)} \end{array}$$

**The meaning of an average.** We are all accustomed to think of the mean (average) as the sum of a series of scores divided by the number of scores in the series. But just how does that mean *represent* the whole series? The answer may be appreciated best by means of an analogy.

Let us suppose that we have three tubes containing colored liquid which stands at heights of 12, 9, and 18 inches in the three tubes, as shown at  $A$  in Figure 1 (page 8). If the tubes are connected and the liquid allowed to flow freely between them, it will seek a level and stand at the same height in the three tubes. What height will this be?

<sup>1</sup> There are certain special terms, like *geometric mean*, which refer to quantities calculated in a way different from the one described above, but the reader need not be concerned with these now.

<sup>2</sup> Note the Greek alphabet on page 290.

## 8 Statistical Method in Educational Measurement

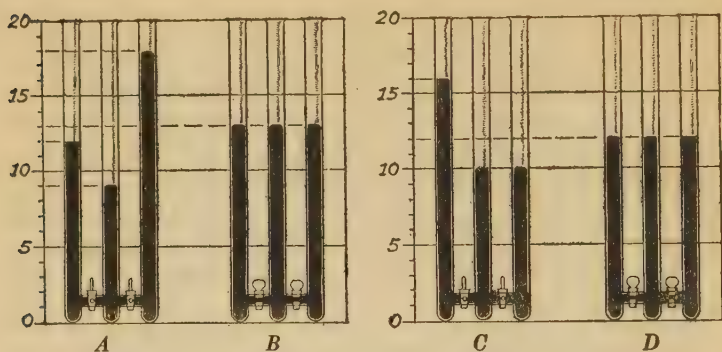


FIG. 1. Illustrating the finding of a mean (average).

To answer this, we might find first how many inches of water there were all together in the three tubes. Adding 12, 9, and 18 gives 39 as the number of inches of water in all, and dividing 39 by 3 gives 13 as the number of inches of water in each tube when at the same level.

What we have done, of course, is to find the mean or average of 12, 9, and 18, which is 13. This analogy furnishes us a new conception of the meaning of a mean or average. The mean or average of a series of numbers may be thought of as the "dead level" that these numbers would assume if they were all made equal.

This conception enables us to use several short methods of finding a mean or average. We can use any method that will give us the "dead level" at which the numbers would stand if made equal. For example, what is the mean of 16, 10, and 10? Let us think of these numbers as representing the heights of liquid in tubes as shown at *C* in Figure 1. We can see at a glance that there is just enough liquid above the level of 10 inches in the first tube (16 inches — 10 inches = 6 inches) to fill up the three tubes 2 inches each above 10 inches, as shown at *D*. The "dead level" of the liquid

will then be at 12 inches. Therefore, the mean or average of 16, 10, and 10 is 12. We may say that 12 is the *central tendency* of the three measures.

Going back to our first illustration (at *A* in Figure 1), let us see how much liquid there is above the level of 9 inches that can be divided equally among the three tubes. There is an excess of 3 inches of liquid above 9 inches in the first tube, an excess of 0 inches above 9 in the second tube, and an excess of 9 inches above 9 in the third tube. Adding these excesses, 3, 0, and 9, gives 12 inches of excess to be divided equally among the three tubes. This is 4 inches to add to 9 inches in each tube, making 13 inches in each tube.

We might have found the mean of 12, 9, and 18 by this method of excesses as follows :

$$\begin{array}{rcl}
 12 & = & 9 + 3 \\
 9 & = & 9 + 0 \\
 18 & = & 9 + 9 \\
 & & \hline
 & & 12 \qquad 12 \div 3 = 4 \\
 & & 9 + 4 = 13 \text{ (the mean)}
 \end{array}$$

The 3 in the first line is the excess of 12 over 9, the 0 is the excess of 9 over 9, and the 9 is the excess of 18 over 9. The total of the excesses is 12, and the average excess over 9 is 4.  $9 + 4 = 13$ . Therefore 13 is the mean or average of the original numbers.

We may state the method used above in the form of a rule, as follows: To find the mean or average of a series of values, (1) find the smallest of those values; (2) find the amount of excess of each of the values over this smallest one; (3) find the average of these excesses, counting zero excesses with the others; and (4) add this average excess to the smallest value. The result is the mean or average of the original series of values.

*Problem:* Apply the rule to find the mean of 62, 65, 69, 72, and 76. *Solution:* The excesses of these numbers over 62,

## 10 *Statistical Method in Educational Measurement*

the smallest, are, respectively, 0, 3, 7, 10, and 14. The average of these five excesses is  $6\frac{4}{5}$ . The mean of the five original numbers is, therefore,  $62 + 6\frac{4}{5}$ , or  $68\frac{4}{5}$ .

There is no necessity for taking the smallest value of a series from which to find the "excesses." In our first illustration, for example, we might have found how much liquid there was in the three tubes above 8 inches. The excesses would then have been, respectively, 4, 1, and 10. These would make a total of 15 inches of liquid to divide among the three tubes, or 5 inches to add to 8 inches in each. This, you will note, gives us the same height, 13 inches, as the "dead level" or mean of the three values.

We might find the mean of 62, 65, 69, 72, and 76 in a similar way. We might take the level 60 from which to figure excesses, since it is easier to subtract 60 than 62. The excesses of the five values would then be 2, 5, 9, 12, and 16, of which the mean is  $8\frac{4}{5}$ . This added to 60 gives  $68\frac{4}{5}$ , as before.

We may state our more general method in the form of a rule, as follows:

**Rule for short method of finding the mean.** To find the mean or average of a series of values, (1) choose as a starting point some convenient value below the values to be averaged; (2) find the average of the excesses of all the values of the series above this chosen value; and (3) add this average excess to the chosen value. The result is the mean or average of the series of values.

Further methods of finding the mean are given on pages 37 and 39.

**EXERCISE 1.** Use the rule just given to find the mean of 84, 89, 96, 81, 95.

**EXERCISE 2.** Try another starting point in Exercise 1 and see if you get the same result as before.

**EXERCISE 3.** Check the results of Exercises 1 and 2 by finding the sum of the five numbers given and dividing by 5.

EXERCISE 4. Use the rule given above to find the mean of each of the following sets of numbers. Use two starting points in each case and see if you get the same result with both. Check the result in each case by finding the average in the usual way (as in Exercise 3).

(a) 10, 15, 12, 12, 11, 16, 13, 16, 14, 10.

(b) 29, 26, 25, 25, 28, 22, 26, 24, 23, 28.

(c) <sup>1</sup> 50, 55, 48, 58, 64, 61, 70, 62, 58, 52.

**The median.** A measure of central tendency which is used more often perhaps than the mean is one called the *median*. Median means "middle." A median score is a middle score as to size; that is, the median score of Grade 5B is the middle score when the scores are arranged in the order of size. The median score is sometimes called the *mid-score*.<sup>2</sup>

Thus, suppose 13 pupils make the following scores :

3, 7, 9, 8, 5, 4, 5, 7, 6, 7, 4, 6, 5.

If these are arranged in order of size, they appear as follows :

3, 4, 4, 5, 5, 5, 6, 6, 7, 7, 7, 8, 9.

M

The middle score as shown by the letter M is 6; that is, the median of these 13 scores is 6.

Let us suppose there had been 14 pupils and the 14th score had been 8. When arranged in order, these appear as follows :

3, 4, 4, 5, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9.

M

The middle of this series is between two scores, the 7th and the 8th, but as they are both 6's, we can call the "median score" 6.

<sup>1</sup> If you wish, you may take 50 as the starting point, and in the case of 48 (the third number), let the "excess" be - 2. It is not necessary to take a starting point below the lowest number. Any numbers below the starting point are merely given *minus* or negative "excesses." In adding the excesses, the negative excesses are subtracted.

<sup>2</sup> This is considered preferable by some — the term *median* to be reserved for a *point* in the middle of the distribution.

## 12 Statistical Method in Educational Measurement

If the 14th score were 5 instead of 8, the series, arranged in order, would appear as follows :

3, 4, 4, 5, 5, 5, 5, 6, 6, 7, 7, 7, 8, 9.

### M

In this case the middle of the series would be between a score of 5 and a score of 6 ; so we could call the median score either 5 or 6, or we could “split the difference” and call it  $5\frac{1}{2}$ . The latter is the more common method.

**Formula for finding the median score.** When there are 13 scores in a series, the middle one is the 7th. Suppose there were 29 scores. Which one (in order of size) would the middle one be? State in words how to tell which is the middle one when there is an odd number of scores in the series. The next paragraph gives the answer.

Whenever the number of scores is odd, the middle one is found by adding 1 to the number of scores and dividing by 2. If there are 29 scores, the middle one will be the 15th.  $(29 + 1) \div 2 = 15$ .

Boiled down to symbols (in which  $N$  stands for the number of scores as before),

Median score = the  $\left(\frac{N + 1}{2}\right)$ th score in order of size.  
(Formula 2, for the median)

*Problem:* Find by means of this formula which score will be the median of a series of 47 scores as in the case of Grade 5B above. *Solution:* When  $N = 47$ , the median score is the  $\left(\frac{47 + 1}{2}\right)$ th score; that is, the 24th score. *Caution:* Do not say the median is 24; it is merely the 24th *score* in order of size, whatever that score may be.

When  $N$  is 14, the median is taken as the value halfway between the 7th and the 8th scores in order of size. If we apply the above formula to this case, we get :

$$\text{Median score} = \left( \frac{14 + 1}{2} \right) \text{th score} = \text{the } 7\frac{1}{2} \text{th score.}$$

In such a case, if we merely interpret  $7\frac{1}{2}$  as meaning the value halfway between the 7th and the 8th scores, which is just where the median is in this case, then Formula 2 above may be considered as correct for finding which score is the median, no matter whether  $N$  is odd or even.<sup>1</sup>

*Problem:* Apply Formula 2, above, to the case of Grade 5A, in which there are 36 pupils, and find which score is the median. *Answer:* When  $N = 36$ , the median score is half-way between the 18th and the 19th cases.

**Finding the median by sorting.** To attempt to search through a series of scores, such as those given in Table 1, to find first the lowest score, then the next to the lowest, etc., would be not only time-consuming, but also very likely to result in error. One excellent way to find the median is to sort the papers themselves in order of score. If this method is used, it is convenient to have the scores written in the margin of the paper, preferably in the upper left-hand corner. The papers may then be laid down one at a time, letting them overlap but always keeping the scores in sight, each successive paper being slipped into its correct position, so that when the last paper is inserted, they will all be in order of scores. The median score may then be found easily by counting.

A second method of finding the median involves the making of a distribution of the scores. It is described on the second page following.

<sup>1</sup> Some books do not give the above formula, but instead consider the median as a point on either side of which  $\frac{N}{2}$  cases fall. Do not be confused by the apparent discrepancy between this expression and the formula given above. The result is the same if properly interpreted. The  $\frac{N}{2}$  rule, however, applies principally to grouped distributions and is discussed in Chapter IV.

**Making a distribution.** For general purposes of interpreting the scores of a group it is desirable to tabulate these in a manner shown in Table 2. In order to make such a table for the scores of Grade 5A, you should begin by writing numbers in a row consecutively from a little below the lowest score that is noticed in the series to a little above the highest score that is noticed.<sup>1</sup> For each score that appears in the

TABLE 2  
ILLUSTRATING THE METHOD OF MAKING A DISTRIBUTION

GRADE	SCORES																NUM. CASES	MED. SCORE
	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
5A					///	/// 		///									36	
5B				///		/// 		/// 	/// 	///							47	
6A																		
6B																		
7A																		
7B																		

series make a mark under the corresponding number in the row just written. Thus, the lowest score that is noticed in the 5A series is 4 and the highest 15. So you should write the numbers from, say, 3 to 16 in order, as shown in Table 2.

By making a mark for each score in Table 1 under the corresponding number in the first row in Table 2 you would have

<sup>1</sup> When the number of scores is large, it is difficult to find the very lowest and very highest by glancing over them, and it is safest to allow a little at each end of the scale.

a row of marks similar to that appearing opposite 5A. This operation is called *distributing* the scores, or making a *distribution*. There are, of course, two distributions in Table 2, one for Grade 5A and one for Grade 5B. The two marks under the score of 4 indicate that the *frequency* of scores of 4 in Grade 5A is 2. The frequency of scores of 5 in Grade 5A is 4; etc.

**Range of score.** There are several advantages in making distributions of scores such as those in Table 2. First, the range of scores of each distribution is shown at a glance. That is, we can see at once that the scores of Grade 5A range from 4 to 15 and those of Grade 5B range from 5 to 16. Each grade therefore has a range of score of 11 points.

**Overlapping.** We can see at a glance also that there is a very great overlapping of scores. That is, there are many scores of pupils in Grade 5A (low fifth in this case) that are higher than those of some pupils in Grade 5B. It may be seen also that the difference between pupils in the same grade is far greater than the difference between grades.

**Finding the median from a distribution.** The chief advantage of a distribution, however, is the ease with which the median may be found from it. Thus, we have already found that the median score of Grade 5A, with 36 cases, is halfway between the 18th and the 19th scores in order of size. We have now only to count from either end toward the middle and find the 18th and the 19th scores. They will be the same two scores no matter from which end we count. In this case (Grade 5A) they are both 8. The median score is therefore 8. Enter this in Table 2.

**Problem:** Find the median score of Grade 5B. **Solution:** Since there are 47 cases, the median is the  $\left(\frac{47+1}{2}\right)$ th or 24th case. Counting from either end, the 24th score is 10.

TABLE 3

SCORES OF GRADES 6A TO 7B IN THE ARITHMETIC REASONING TEST

GRADE	SCORES											
6A	14,	14,	12,	8,	15,	8,	9,	13,	12,	11,	6,	12,
	14,	4,	12,	7,	14,	10,	14,	4,	9,	13,	13,	18,
	7,	7,	13,	10,	12,	11,	9,	15,	10,	5,	12,	3,
	12,	10,	6,	14,	12,	14,	10,	8,				
6B	12,	9,	15,	10,	4,	13,	10,	8,	8,	10,	12,	3,
	13,	11,	8,	10,	9,	15,	5,	16,	6,	11,	11,	10,
	12,	11,	10,	12,	8,	13,	11,	11,	11,	11,	8,	10,
	8,	12,	14,	3,	18,	13,	14,	12,				
7A	15,	13,	11,	8,	13,	11,	10,	10,	8,	9,	12,	15,
	8,	11,	11,	10,	14,	12,	7,	9,	14,	13,	14,	9,
	5,	8,	12,	11,	11,	15,	5,	11,	10,	13,	15,	11,
	13,	6,	13,	5,	6,							
7B	9,	10,	11,	14,	6,	9,	13,	15,	12,	13,	14,	10,
	12,	10,	13,	10,	10,	10,	14,	7,	14,	13,	13,	8,
	11,	10,	12,	16,	3,	10,	14,	13,	14,	9,	15,	8,

Therefore the median score of Grade 5B is 10. Enter this in Table 2.

**EXERCISE 5.** Continue Table 2, or make a table similar to it, showing the distributions of scores of Grades 6A to 7B in the arithmetic reasoning test as given in Table 3. Find the median score of each grade.

**Special cases.** The median of the scores 3, 4, 5, 7, 8, 9 is called 6, since it is halfway between the 3d and 4th scores; that is, halfway between 5 and 7. Similarly the median of the scores 2, 3, 4, 7, 8, 9 is  $5\frac{1}{2}$ , since it is halfway between 4 and 7. The same method is used for any interval.

Various refinements sometimes used in the finding of a median more precisely are discussed in later chapters.

TABLE 4

SHOWING THE METHOD OF FINDING THE MEAN OF A DISTRIBUTION OF SCORES

SCORE	FREQUENCY	PRODUCT
15	1	15
14	0	0
13	1	13
12	3	36
11	3	33
10	5	50
9	4	36
8	7	56
7	5	35
6	1	6
5	4	20
4	2	8

$$\begin{array}{r} 36 \overline{) 308} (8.6 \\ \underline{288} \\ 200 \end{array}$$

**Finding the mean score from a distribution.** Let us suppose that we wish to compare Grades 5A and 5B in arithmetical ability by comparing the mean (average) scores of the two grades in the arithmetic test. It would be somewhat tedious to copy all the scores of Grade 5A in a column and add them, and to do the same for the scores of Grade 5B. A shorter method is to make a distribution of scores for each grade, as shown in Table 2, and to find the mean from this distribution.

Thus, to get the sum of the scores of Grade 5A, we need not add 15 once, 13 once, 12 three times, 11 three times, 10 five times, etc. It is easier to multiply each amount of score by the frequency of scores of that amount and add the products, as shown in Table 4. Thus the 3 scores of 12 make 36, the 3 scores of 11 make 33, etc. The sum of these products is 308. Dividing this sum by 36, the number of scores, gives 8.6 as the mean score of Grade 5A.

We might have used the short method previously learned for finding the mean, of course, although in this case it is hardly worth while. Thus, we might have found the amounts by which each score deviated from 4, say, and multiplied these amounts by the frequencies, as shown in Table 5.

TABLE 5

SHOWING THE METHOD OF FINDING THE MEAN OF A DISTRIBUTION OF SCORES BY THE SHORT METHOD

SCORE	FREQUENCY	DEVIATION FROM 4	PRODUCT (Frequency X Deviation)
15	1	11	11
14	0	10	0
13	1	9	9
12	3	8	24
11	3	7	21
10	5	6	30
9	4	5	20
8	7	4	28
7	5	3	15
6	1	2	2
5	4	1	4
4	2	0	0

36)164(4.6

144

200

$$4 + 4.6 = 8.6 = \text{mean}$$

The mean found by this method is 8.6, which, of course, must be exactly the same as found by the first method (Table 4). This method is used very frequently when finding coefficients of correlation.

**The mode.** We have discussed two measures of central tendency thus far: the mean and the median. There is a third measure sometimes used to denote the central tendency. This is the *mode*. The mode is the score that occurs most frequently. Thus in Grade 5A (see Table 2) the score of 8

occurred most frequently. The mode or modal score of Grade 5A is therefore 8. There are two scores in Grade 5B that occurred most frequently. The scores of 8 and 11 each occurred eight times. When there are two modes with scores between them in this way, the distribution is spoken of as *bimodal*.<sup>1</sup>

**Summary.** There are three common measures of central tendency, the mean, the median, and the mode.

The *mean* is the sum of the scores divided by the number of scores.

The *median* is the middle score or middle point.

The *mode* is the score that occurs most frequently. (It is the fashion.)

### QUESTIONS

1. Which one of the three measures of central tendency seems to you to be the best representative measure of a whole group? Why?

2. Which seems the least representative? Why?

3. Does the median or mean of a distribution of scores completely represent the distribution? If not, what aspects of a distribution are not exhibited by the median or mean? That is, can you show ways in which two distributions may differ in general and yet have the same mean or median?

<sup>1</sup> The term *bimodal* generally refers to a distribution in which there is a distinct tendency for high frequencies in two different parts of the scale, with lower frequencies between them. It is probable that the two modes are accidental in this case, and that on a second test only one mode would appear. For this reason the mode is not a satisfactory measure of central tendency when there are few cases.

## CHAPTER THREE

### REPRESENTING THE WHOLE OF A DISTRIBUTION

**Comparison of individuals within a group.** It is often desirable not only to compare groups such as Grades 5A and 5B but also to compare an individual with the group of which he is a member. Thus, for example, we might wish to find out how an individual in Grade 5A who made a score of 6 stood with respect to the group.

We have found the median score of the group to be 8; so we can say at once that the score of 6 is below the median. But how much below the median? In what quarter of the scores will the score of 6 fall — the third quarter or the fourth quarter? (The highest quarter is called the first quarter.) You will see that it is not evident at a glance at the distribution of scores of Grade 5A in Table 2 whether the score of 6 falls in the third or the fourth quarter.<sup>1</sup>

Perhaps the simplest way to make it possible to see at a glance whether the score of 6 falls within the third or the fourth quarter is to write the scores in a column, writing each number as many times as the score appears, as shown in the first column of Table 6. Since there are 36 scores, we can divide these into four quarters of 9 scores each as shown, and it can then be seen at a glance that the score of 6 falls in the lowest or fourth quarter.

**Rank order.** We may state the relation of the score of 6 to the whole group of scores still more definitely by finding the rank of that score among the scores of the class. Thus, having arranged the scores in order of size, in the first column of Table 6, we may number them consecutively from 1 to 36,

<sup>1</sup> This may seem contrary to the custom of counting from left to right, but statisticians have established the rule of calling the highest quarter the first quarter and it would be confusing to change.

TABLE 6

SHOWING THE RANK ORDER OF SCORES OF PUPILS IN THE 5A GRADE  
IN THE ARITHMETIC REASONING TEST

SCORE	RANK	SCORE	CORRESPONDING RANKS	MEDIAN RANK ASSIGNED
15	36	15	36	36
13	35	13	35	35
12	34	12	34	33
12	33		33	
12	32		32	
11	31	11	31	30
11	30		30	
11	29		29	
10	28	10	28	26
10	27		27	
10	26		26	
10	25		25	
10	24		24	
9	23	9	23	21 $\frac{1}{2}$
9	22		22	
9	21		21	
9	20		20	
8	19	8	19	16
8	18		18	
8	17		17	
8	16		16	
8	15		15	
8	14		14	
8	13	7	13	10
7	12		12	
7	11		11	
7	10		10	
7	9		9	
7	8	6	8	7
6	7		7	
5	6		6	
5	5	5	5	4 $\frac{1}{2}$
5	4		4	
5	3		3	
4	2	4	2	1 $\frac{1}{2}$
4	1		1	

## 22 Statistical Method in Educational Measurement

beginning with the lowest score,<sup>1</sup> as shown. The score of 6 is seen to be No. 7 in the group; so we may say that the pupil who made the score of 6 has a rank of 7 among the 36 members of his class, in score.

The pupil who made a score of 15 has a rank of 36 (a rank of 36 among 36; therefore the highest). There are three pupils who made scores of 12. They occupy ranks of 34, 33, and 32, but since it would be unjust to say arbitrarily that a particular one of these three pupils ranks above the other two, we assign each the median of these three ranks, which is 33, as shown in the column at the right in Table 6.

Similarly, four pupils made scores of 9, and in order to give them all the same rank, we shall assign each the median of ranks 20, 21, 22, and 23, which is a rank of  $21\frac{1}{2}$ .

Whenever the number of pupils making the same score happens to be even, the rank assigned will contain the fraction  $\frac{1}{2}$ . When the number is odd, the rank will be an integer (whole number).

*Problem:* Find the rank of each score of Grade 5B in Table 1. *Solution:* Instead of writing the 47 scores of Grade 5B in order, we may refer to the distribution in Table 2. From this table we see that there is one score of 5 which is the lowest score; therefore the rank of the score of 5 is 1. There are five scores of 6; therefore the rank of a score of 6 is the median of the next five ranks, 2, 3, 4, 5, and 6, which is 4, etc., as shown in Table 7.

**EXERCISE 6.** Find the rank of each score of Grade 6A in the arithmetic reasoning test, using the method shown in Table 6. The distribution of scores was found in Exercise 5, page 16.

<sup>1</sup> Ordinarily, of course, if an individual is said to rank first in a group, we understand this to mean that he ranks highest. As leading up to the discussion of percentile ranks, however, in which a high rank is represented by a high number, it is best here to begin numbering the individuals at the bottom. For all practical purposes this method of ranking works exactly as well as the other.

TABLE 7

SHOWING THE METHOD OF FINDING THE RANK OF EACH SCORE  
OF A GROUP FROM THE DISTRIBUTION OF SCORES

SCORE	INDIVIDUAL RANKS	MEDIAN RANK ASSIGNED
16	47	47
15		
14	45   46	$45\frac{1}{2}$
13	41 42   43 44	$42\frac{1}{2}$
12	36 37 38 39 40	38
11	28 29 30 31   32 33 34 35	$31\frac{1}{2}$
10	21 22 23 24   25 26 27	24
9	17 18   19 20	$18\frac{1}{2}$
8	9 10 11 12   13 14 15 16	$12\frac{1}{2}$
7	7   8	$7\frac{1}{2}$
6	2 3 4 5 6	4
5	1	1

EXERCISE 7. Find the rank of each score in Grade 6B in the arithmetic reasoning test, using the method shown in Table 7.

EXERCISE 8. How does a score of 10 compare in rank in Grades 6A and 6B?

**Comparison of pupils of different groups.** For purposes of comparing the score of any individual with the scores of the group of which he is a member, it is quite sufficient to find the rank of the individual in the group. But if we wish to compare the relative rank of a pupil in one group with that of another pupil in another group, we cannot use rank orders conveniently. For example, how does a pupil in Grade 5A making a score of 11 compare in relative rank with a pupil in Grade 5B making a score of 12?

According to Table 6 a score of 11 has a rank of 30 in Grade 5A. According to Table 7 a score of 12 has a rank of 38 in Grade 5B. Does this mean that a score of 12 ranks relatively higher in Grade 5B than a score of 11 does in Grade 5A? No, because in Grade 5A a score of 11 ranks 30 among 36, while in Grade 5B a score of 12 ranks only 38 among 47.

**The percentile rank.** In order to make a direct comparison between these relative ranks we must reduce them to the same terms — to a common basis. For this purpose a rank calculated on the basis of 100 is used. This is called the *percentile rank*.

The percentile rank of an individual in a group is sometimes rather loosely defined as being the per cent of pupils in the group whom he exceeds in score. Thus a percentile rank of 75 would mean, according to this conception, that the individual would exceed 75 per cent of the individuals in the group in score. This is a good, simple, working conception for general use, but if fine distinctions are needed it is necessary to refine the definition a little.

You can see that the median pupil in a group of 25 would exceed 12 in score, and according to the definition given above, this pupil would have a percentile rank of  $\frac{12}{25}$  of 100, or 48. The lowest pupil, exceeding none, would have a percentile rank of 0, while the highest, exceeding 24 of the 25, would have a percentile rank of  $\frac{24}{25}$  of 100, or 96. Now for convenience in interpretation we must define percentile rank so that the median pupil will always have a percentile rank of just 50 and so that the percentile rank of the highest and the lowest will be equally above and below 50.

*Let us define the percentile rank of an individual, then, as the percentage of a very, very large group of individuals<sup>1</sup> (corresponding exactly to the group in question) whom the individual would exceed in score.*

Now the most probable position that the lowest pupil in a group of 25 would have in a very, very large group of the same sort (say a group of a million) is in the middle of the lowest 25th. In other words, he would most probably exceed  $\frac{1}{50}$ , or 2 per cent, of the group. His percentile rank in the group of 25 according to our new definition would be 2. Similarly the

<sup>1</sup> Theoretically an infinite number.

most probable position that the highest pupil in the group of 25 would have in the group of a million would be in the middle of the upper 25th; that is, he would most probably exceed  $\frac{49}{50}$  of the group. His percentile rank would be, therefore, 98. The median pupil of the group of 25 would most probably be in the middle of the middle 25th of the million pupils or, in other words, he would most probably exceed just 50 per cent of the million pupils and have, therefore, a percentile rank of 50.

According to our more precise definition of a percentile rank, then, the median of any group will always have a percentile rank of just 50, and the highest and the lowest pupils will have percentile ranks equally above and below 50. It will be found also that pupils equally distant above and below the median in rank will have percentile ranks also equally above and below 50.

The precise definition of a percentile rank given above will serve to establish its meaning in our minds, but for working purposes it is simpler to use a more concrete conception. This may be done by thinking of the individuals composing the group as standing shoulder to shoulder in a line and occupying a given space which is thought of as 100 units long and graduated from 0 to 100. The position on the scale to which any individual's nose points is his percentile rank.

If there were 23 individuals in the group, it would be as if we placed 23 circles in a row as shown in Figure 2 and scaled the space they occupied from 0 to 100 as shown. Then the percentile rank of the lowest individual, according to our new concept, is the point on the scale opposite the *center* of

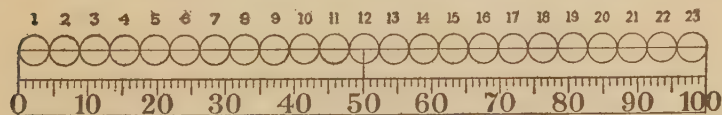


FIG. 2. Showing a convenient conception of the method of finding the percentile rank of an individual in a group.

the first circle. This is, of course,  $\frac{1}{2}$  of  $\frac{1}{23}$  of 100, or 2.17. So we may say that the lowest individual in a group of 23 has a percentile rank of 2.17, or roughly 2, in that group. The second individual has a percentile rank equal to  $1\frac{1}{2}$  times  $\frac{1}{23}$  of 100, which equals 6.52, or roughly 7, etc. The median or 12th individual has a percentile rank of exactly 50 by this method, and the percentile rank of the highest individual is just as much below 100 as the rank of the lowest is above 0.

**The precise rule for finding a percentile rank.**<sup>1</sup> By the precise method just described the percentile rank of the 3d individual would be  $2\frac{1}{2} \times \frac{1}{23} \times 100$ ; etc. And in general, to find the percentile rank of an individual having any given rank among the 23, we should multiply  $\frac{1}{23}$  of 100 by the rank less  $\frac{1}{2}$ .

**Formula for the precise determination of a percentile rank.** For convenience in using this rather complicated rule, let us convert it into a formula. The first step in making a formula is to write the rule in arithmetic form, thus:

$$\text{Percentile rank} = \left( \text{rank} - \frac{1}{2} \right) \times \frac{1}{\text{number of cases}} \times 100.$$

We may now condense it still more by using symbols. Thus, let *P.R.* stand for percentile rank, let *N* stand for the number of cases, and *R* stand for the rank of the given individual among the *N* cases. Then the formula can be expressed:

$$P.R. = \left( R - \frac{1}{2} \right) \times \frac{1}{N} \times 100.$$

If you are familiar with algebraic manipulations, you will see that this is the same as

$$P.R. = \left( \frac{2R - 1}{2} \right) \times \frac{1}{N} \times 100,$$

or

$$P.R. = \left( \frac{2R - 1}{2N} \right) \times 100. \quad (\text{Formula 3})$$

<sup>1</sup> A more simple and convenient rule is given below (page 28).

This is a fairly simple formula. For example, let us return to the comparison of the relative ranks of 30 among 36 and 38 among 47, discussed on page 23.

To find the percentile rank of the first pupil (whose rank is 30 among a group of 36) we substitute 30 for  $R$  and 36 for  $N$  in the formula and solve; thus:

$$\begin{aligned} P.R. &= \frac{2 \times 30 - 1}{2 \times 36} \times 100 \\ &= \frac{59}{72} \times 100 \\ &= 82. \end{aligned}$$

Therefore the first pupil's percentile rank is 82. To find the second pupil's percentile rank we should substitute 38 for  $R$  and 47 for  $N$  in the same formula. In that case

$$P.R. = \frac{2 \times 38 - 1}{2 \times 47} \times 100 = 80. \text{ The second pupil's per-}$$

centile rank, therefore, is 80, and we can compare this directly with the other pupil's percentile rank of 82 and see that a rank of 38 among 47 is relatively less than a rank of 30 among 36.

You will see that in no actual case can the true percentile rank of an individual be 100 or 0. The greater the number of cases, the nearer to 100 and 0 the percentile ranks of the highest and lowest individuals become, as shown in Table 8.

TABLE 8

NUMBER OF CASES	10	25	50	100	500	5000	Infinite
P.R. OF HIGHEST	95	98	99	99.5	99.9	99.99	100.
P.R. OF LOWEST	5	2	1	0.5	0.1	0.01	0.

An approximate method of finding a percentile rank. For ordinary purposes it is hardly necessary to use the precise

method just described for finding a percentile rank. A simple rule, which works well for practical use, is to consider a pupil's position among the members of the group, *not counting himself* — in other words, to find the percentage of pupils in the group, not counting himself, whom he exceeds in score. Thus, if there are 23 pupils in a group, the median pupil exceeds 11 out of the 22 *others*. This would be 50 per cent; so his percentile rank according to this rule is 50, the same as it would be if we used the method just described. The pupil having a rank of 4 in 23, however, exceeds 3 of the 22 *others*; so his percentile rank by this method is  $\frac{3}{22} \times 100$ , or 13.6. It will be seen that by this approximate method, no matter how many there are in the group the pupil who ranks highest gets a percentile rank of 100 and the pupil who ranks lowest gets a percentile rank of 0. Moreover, by this method the median score always has a percentile rank of exactly 50, since the middle score always exceeds exactly 50 per cent of the other scores.

*The simple rule for finding the percentile rank of a pupil in a group, therefore, is to find the percentage of the group, not counting himself, whom he exceeds in score.*

The greatest difference in percentile ranks found by the two methods is at the extremes. Thus, where we got a percentile rank of 0 for the lowest case in our illustration by the approximate method, we got  $\frac{1}{2}$  of  $\frac{1}{23}$  of 100 by the precise rule. For practical purposes, however, this difference is not serious. In other words, although the precise rule is more nearly correct, the approximate method may as well be used for practical purposes.

**Interpreting the rule.** In calculating percentile ranks by either method, care must be taken to interpret the rule properly when several pupils make the same score. Five pupils in Grade 5A made scores of 10. The rank of the middle one of the five is 26. Each is considered as if he were

the middle one. That is, he is considered as if he *exceeded* two of the other four and as if he were exceeded by the other two. Therefore, be sure to count one half of the other pupils (those making the same score as the pupil in question) as if they were exceeded by that pupil.

If there had been six pupils who made scores of 10, then in finding the percentile rank of any one of these six pupils we must assume that he exceeds one half of the other five; that is, we must assume in making our calculation that he exceeds  $2\frac{1}{2}$  of those who made scores of 10 as well as those who made scores below 10.

*Problem:* By the simple rule, find the percentile rank of a pupil in Grade 5A making a score of 10. (See Table 6.)

*Solution:* We assume the pupil in question to be the middle one of the five making scores of 10. He is therefore considered as exceeding 25 pupils in score. There are 35 in the grade, not counting himself. 25 is about 72 per cent of 35. Therefore we should call his percentile rank in the grade 72.

*Problem:* By the simple rule, find the percentile rank of a pupil in Grade 5A making a score of 9. *Solution:* The pupil in question is assumed to exceed  $1\frac{1}{2}$  of the other three making scores of 9. He is therefore considered as exceeding  $20\frac{1}{2}$  pupils in score.  $20\frac{1}{2}$  is about 60 per cent of 35. His percentile rank is therefore 60.

*Problem:* By the simple rule, find the percentile rank among the pupils of Grade 5A that a pupil would have who made a score of 14. *Solution:* There is no score of 14 in the Grade 5A distribution, but if we added a score of 14, it would exceed 35 of the 36 other scores. 35 is about 98 per cent of 36. Therefore we may say that a pupil making a score of 14 would have a percentile rank of 98 among the pupils of Grade 5A. Percentile ranks corresponding to other scores are shown in Table 9.

TABLE 9

SCORE		4	5	6	7	8	9	10	11	12	13	14	15	16
PERCENTILE RANK	5A	4	12	19	28	44	60	72	83	92	97	98	100	
	5B													

EXERCISE 9. Find by the simple rule and record in Table 9 the percentile ranks corresponding to scores from 5 to 16 in Grade 5B.

**The bar graph.** We may show the distribution of scores still more plainly than in Table 6 by the graphic representation shown in Figure 3. This figure is called a *bar graph*. In it each score in Grade 5A is represented by a bar, the height of which corresponds to the amount of the score. Thus, the first two bars at the left represent the two scores of 4, since they are each 4 units high according to the scale of scores at the left; the next four bars represent the four scores of 5; the next bar represents the score of 6, which is now plainly seen to fall within the fourth quarter, etc.

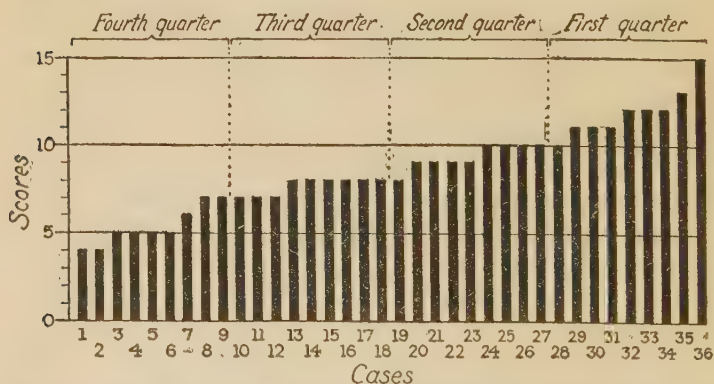


FIG. 3. A bar graph representing the distribution of scores of 36 pupils in Grade 5A in the arithmetic reasoning test.

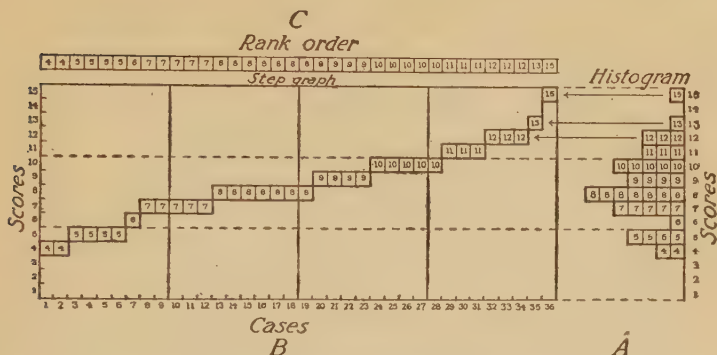


FIG. 4. A step graph and histogram representing the scores of 36 pupils in Grade 5A in the arithmetic reasoning test.

**The histogram.** Using a scale of scores similar to that at the left in Figure 3, we may represent the distribution of scores graphically in a different way, as shown at *A* in Figure 4, by representing the two scores of 4 by two squares opposite 4 in the scale, representing the four scores of 5 by four squares opposite 5 in the scale, etc. The area occupied by these thirty-six squares is called a *histogram*. A histogram is often drawn so as to be composed of rectangles only, and to rest upon a horizontal base, as shown at *A* in Figure 5. The same histogram is shown at *B* with outline only. When individual squares are not shown in a histogram, as in the case of those in Figure 5, it is customary to provide a scale of

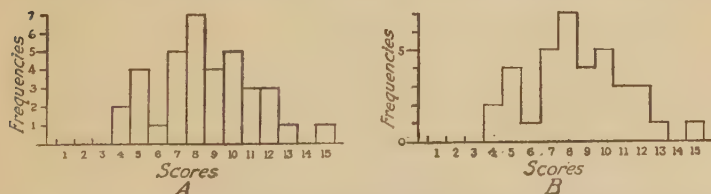


FIG. 5. Histograms representing the distribution of scores of Grade 5A in the arithmetic reasoning test, showing the use of rectangles and of outline only.

frequencies, as shown, by which the heights of the rectangles may be gauged.

**The step graph.** At *B* in Figure 4 is shown a graph that is in the nature of a combination of the histogram at *A*, the rank order at *C*, and the bar graph in Figure 3. It is like the histogram and bar graph in that each score is represented by a square the height of which above the base line corresponds to the amount of the score and in that the frequency of each score is easily seen. It is like the rank order and bar graph in that the scores are arranged continuously so that it is possible to determine very easily the relative position of any score, as for example the score of 6; and by dividing the horizontal distance into four equal parts it can be seen at a glance what scores fall within each quarter. It is not possible to divide a histogram into four quarters as handily.

The intimate connection between the histogram and step graph is suggested by the arrows. We might imagine that cubical beads at *A* have been slid along wires to the positions at *B*. It is not necessary, of course, to write in the squares the amounts of score, or even to complete each of the small squares in a step graph.

*Problem:* Draw a step graph similar to that shown in Figure 4 to represent the distribution of score of Grade 5B. Use cross-section paper ruled about five lines to the inch. (See Figure 8.) Begin by laying out a scale of scores at the left and a scale of cases at the bottom. Make a series of rectangles each containing as many squares as there are scores in the corresponding frequency. Divide the horizontal distance into four equal parts by vertical lines. Simplify by showing only the outlines of the rectangles and by omitting amounts of score within the rectangles. *Solution:* The resulting step graph is shown in Figure 6.<sup>1</sup>

<sup>1</sup> It would not be practical to make a step graph even as simple as that in Figure 6, merely in order to find within which quarter of a distribution a given score falls. A simplification of the step graph, which has a very prac-

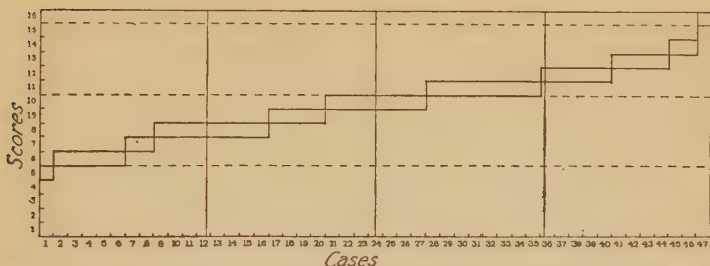


FIG. 6. A step graph representing the distribution of scores of Grade 5B in the arithmetic reasoning test (simplified by not showing separate squares or amounts of score within rectangles).

**Finding the median from a step graph.** It will be seen that since there are 47 cases, the lines dividing the horizontal distance into four equal parts do not mark off whole numbers of scores. We may say there are  $11\frac{3}{4}$  scores in each quarter. The middle line passes through the rectangle representing scores of 10. Therefore we may say that the median score is 10. This, of course, is the same score we found as the median in Chapter I.

**Quartile scores.** The line dividing off the upper quarter of scores passes through the rectangle representing scores of 12. We call this score the *upper-quartile* score of the distribution. The upper-quartile score is called  $Q_3$ . The line dividing off the lower quarter of scores passes through the rectangle representing scores of 8. Therefore we call the score of 8 the *lower-quartile* score. The lower-quartile score is called  $Q_1$ . The three scores, 12, 10, and 8, are therefore  $Q_3$ ,  $M$ , and  $Q_1$ . (We might think of the median as  $Q_2$ .)

In the case of Grade 5A,  $Q_3 = 10$ ,  $M = 8$ , and  $Q_1 = 7$ . If the upper dividing line in Figure 4 had passed between the rectangles representing scores of 10 and 11 (one unit to the

tical value, is explained in a later chapter. The step graph as shown in Figure 6, however, is easiest to understand at first and will help to make clear the more practical graph developed from it.

right), we should say that  $Q_3 = 10\frac{1}{2}$ . If the range of scores is great enough so that a half point is negligible, it is customary to take either one or the other nearest whole number (say the even number) for  $Q_1$  or  $Q_3$ .

The use of  $Q_1$  and  $Q_3$  is discussed in Chapter VIII.

**The line of diagonals.** We can go still further in simplifying the step graph by omitting the rectangles altogether and drawing merely the diagonals. The position of a rectangle, of course, may be completely determined by the position of its diagonal. Figure 7 shows the step graph of Figure 4, with the rectangles replaced by diagonals. The line in Figure 7 we may call a *line of diagonals* or *line graph*. This line of diagonals also shows that there were two scores of 4, four scores of 5, one score of 6, etc., and no score of 14.

*Problem:* Draw a line of diagonals, similar to that shown in Figure 7, to represent the distribution of scores of Grade 5B in the arithmetic reasoning test. Omit rectangles altogether. (See Figure 6.) *Solution:* The resulting graph is shown in Figure 8.

To draw a line of diagonals of this sort, it is necessary merely to locate the points where the diagonals begin and end. Thus the point marking the lower end of the first diagonal must be at the left edge of the graph and on the line marking the lower limit of the space representing the score

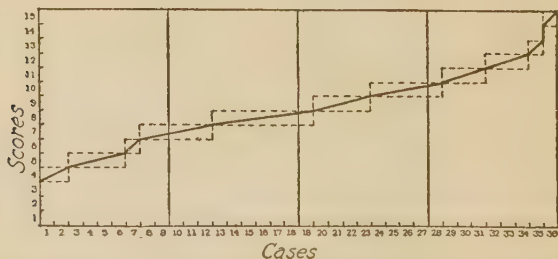


FIG. 7. The step graph of Figure 4 with rectangles replaced by diagonals.

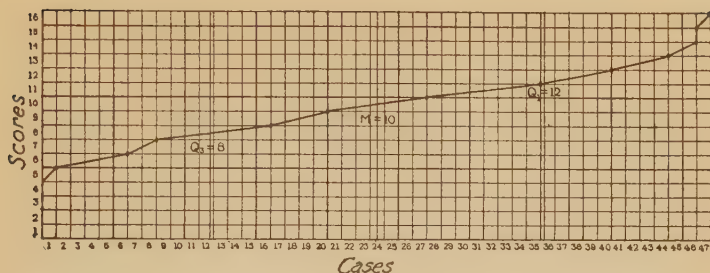


FIG. 8. The line of diagonals representing the distribution of scores of Grade 5B in the arithmetic reasoning test.

of 5, as shown in the scale at the left. The next point must be up one space and over one, since there was *one* score of 5. The next point must be up one and over *five*; the next, up one and over *two*; etc. Each point is located by moving the pencil up one space (the width of the rectangle) and over to the right as many spaces as there are scores in the frequency that is being represented. In the case of score 15, since there is none, we go up one and over none. A vertical "diagonal," then, means no score. The longer the diagonal and the more nearly level it is, the greater the frequency of score.

EXERCISE 10. Using a sheet of cross-section paper (preferably ruled about five lines to the inch), make an enlarged copy of Figure 8.

EXERCISE 11. Draw a line of diagonals similar to that shown in Figure 8 to represent the distribution of scores of Grade 6A. (See Exercise 5, page 16, or Table 3.) Find  $Q_1$ ,  $M$ , and  $Q_3$ .

EXERCISE 12. Draw a histogram similar to that shown at B in Figure 5, to represent the distribution of scores of Grade 6A.

EXERCISE 13. Draw a histogram and a line of diagonals to represent the distribution of scores of Grade 6B. (See Exercise 5.)

In a later chapter we shall see how a line of diagonals may be drawn on a special graph called a Percentile Graph so that the percentile rank of any score may be read at once.

## CHAPTER FOUR

### GROUPED DISTRIBUTIONS

**Need for a new method of dealing with scores.** In all our discussions so far we have dealt with distributions of scores covering only a narrow range, — as, for example, from 4 to 15, — so that the frequency of each score might be considered. When a distribution has a wide range, — as, for example, from 15 to 69, — it is more convenient to use a different method of handling it — one in which scores are tabulated in groups.

In Table 10 are given the scores of Grades 4B, 5A, 5B, and 6A in a mental-ability test.

TABLE 10

GRADE	SCORES											
4B	13,	18,	19,	38,	41,	9,	31,	30,	26,	20,	45,	54,
	20,	7,	30,	18,	10,	19,	9,	44,	16,	26,	22,	25,
	28,	16,	10,	20,	23,	14,	17,	38,	28,	36,	17,	31,
	47,	22,										
5A	23,	35,	37,	45,	27,	16,	31,	42,	28,	35,	54,	34,
	47,	22,	48,	39,	27,	45,	26,	29,	13,	29,	41,	37,
	27,	18,	48,	32,	40,	16,	42,	30,	19,	30,	15,	26,
5B	38,	33,	20,	22,	47,	66,	59,	19,	37,	34,	47,	45,
	32,	18,	48,	28,	27,	45,	57,	43,	41,	47,	44,	50,
	31,	44,	33,	28,	30,	52,	61,	23,	41,	24,	26,	23,
	57,	48,	62,	39,	17,	36,	17,	41,	46,	19,	31,	
6A	58,	69,	30,	26,	60,	47,	53,	46,	57,	52,	40,	56,
	49,	34,	66,	50,	53,	49,	66,	23,	34,	47,	45,	54,
	49,	23,	59,	38,	38,	34,	41,	46,	61,	47,	26,	53,
	30,	24,	66,	39,	59,	45,	50,	41,	35,			

Following the same plan that we used with the scores in the arithmetic test, let us first make a distribution of the scores

in the mental-ability test. If we made a table similar to Table 2 and marked down each score separately, our table would be so long as to be quite unwieldy. It is much more convenient to group the scores in such a way that the number of groups will not exceed twenty and preferably not be much over ten. Suppose we group the scores by fives, bringing together scores from 0 to 4, 5 to 9, etc. These intervals of score are often called *class intervals*. The scores of Grades 4B, 5A, and 5B thus distributed appear as shown in Table 11.

EXERCISE 14. Complete Table 11 by distributing the scores of Grade 6A. (See Table 10 for scores.)

TABLE 11

SHOWING THE DISTRIBUTION OF SCORES OF GRADES 4B, 5A, AND 5B IN THE MENTAL-ABILITY TEST

SCORE INTERVALS	0 TO 4	5 TO 9	10 TO 14	15 TO 19	20 TO 24	25 TO 29	30 TO 34	35 TO 39	40 TO 44	45 TO 49	50 TO 54	55 TO 59	60 TO 64	65 TO 69	70 TO 74	75	No. OF CASES
Grade 4B				 	 												38
Grade 5A						 											36
Grade 5B							 		 								47
Grade 6A																	

**Finding the mean of a grouped distribution.** For ordinary purposes of representing the central tendency of a distribution the median suffices. It is nearly always used, for example, in finding *norms*. Thus the norm (or normal score) for the fifth grade in any test is generally considered as being the median score of a large group of unselected fifth-grade pupils. This will be discussed further in a later chapter.

There are occasions, however, when it is desirable to find the mean score of a distribution. This is necessary, for example, in finding a coefficient of correlation by the usual method.

When we have distributed the scores of a group of pupils, as in Table 11, thereby losing track of the actual scores, we are not able to find the mean in the ordinary way — that is, by adding the scores and dividing by the number of scores. The only way left is to assume that all the scores that fell in a given interval, such as 35–39, have the middle value of that interval and find the mean of these values.

*Problem:* Find the mean score of Grade 5A in the mental-ability test. (See Table 11.) *Solution:* The score that fell in the interval 50–54 we shall assume to be a score of 52. The next five scores we shall assume to be 47, etc. The calculation of the mean is, then, as follows:

1 score of 52	=	52
5 scores of 47	=	235
4 scores of 42	=	168
5 scores of 37	=	185
5 scores of 32	=	160
8 scores of 27	=	216
2 scores of 22	=	44
5 scores of 17	=	85
1 score of 12	=	12
<hr/>		
sum of 36 scores	=	1157
<hr/>		
$1157 \div 36 = 32.14$ (the mean)		

In expressing the mean as 32.14, it is assumed that a precise measure is needed. If such is not the case, we may call the mean 32.

If we go back to the original scores and find the mean in the usual way, this comes out 32.03. This shows that our calculation based upon the grouped distributions of Table 11 was correct within about one third of one per cent.

**A more convenient method of finding the mean (Solution by substitution).** What is the average of 3 ft., 6 ft., 9 ft., 12 ft., and 15 ft.? *Answer:* 9 ft. Suppose we changed these measures to yards. We should then have 1 yd., 2 yd., 3 yd., 4 yd., and 5 yd. What is the average of these five measures? *Answer:* 3 yd. If we translate this back to feet, we have 9 ft., the same answer that we got at first. Bear this idea in mind while reading what follows.

Let us consider another introductory illustration. Imagine a single thermometer graduated in both Fahrenheit and Centigrade scales. The mean of a series of Fahrenheit temperatures can be represented, of course, by a certain point on the thermometer and if we found the mean of those *same temperatures* expressed in Centigrade the mean would be represented necessarily by the same point on the thermometer. We can apply this principle to the calculation of a mean so as to simplify it.

In Figure 9 is shown a line graduated with two scales, *S* and *A*. The *S* scale represents actual score values; the *A* scale is an auxiliary scale. These two scales may be assumed to correspond, just as a Fahrenheit scale and a Centigrade scale correspond. Now the mean of a series of values on the *S* scale may be represented by a certain point on the scale. The mean of the *A* values corresponding to those *S* values will be represented by the same point on the scale, as was explained with reference to a thermometer. Therefore to find the mean of a series of *S* values (score values) we have but to find the mean of the corresponding *A* values (auxiliary values) and convert this into terms of *S* values. Let us take a very simple case as an example.

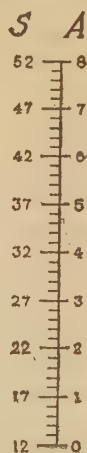


FIG. 9. Showing the method of substituting an auxiliary scale of larger units and smaller numbers for a scale of scores for finding the mean.

*Problem:* Find the mean of the numbers 12, 17, 22, 27, and 27.

*Ordinary solution:* The sum of these numbers is 105. Dividing this by 5 gives 21 as the mean.

*Solution by substitution:* Now let us substitute for these numbers the corresponding *A* values (auxiliary) shown in Figure 9. These are 0, 1, 2, 3, and 3. The sum of these values is 9;  $9 \div 5 = 1.8$ . Now 1 on the *A* scale equals a score of 17. What does .8 of a unit on the *A* scale represent in terms of score? Since 1 unit on the *A* scale represents 5 units of score, .8 on the *A* scale represents .8 of 5, or 4 units. This number added to 17 gives 21 as the mean of the five scores. This is, of course, the same value we got by the ordinary solution.

*Problem:* Find the mean of the scores of Grade 5A by the two methods. *Solution:* The two solutions are shown at the top of the next page. On the left is shown the ordinary solution. On the right is the solution by substitution. For the score of 52 the auxiliary scale value 8 has been substituted. For the score of 47 the auxiliary scale value 7 has been substituted; etc.

By comparing these calculations you will see that in the second we first found the mean of the auxiliary values corresponding to the score values in the first calculation. This mean came out 4.028. What *S* value (score value) corresponds to this *A* value? The point 4 on the *A* scale corresponds to a score of 32. What is the score value of .028 points on the *A* scale? By examining Figure 9 you will see that each unit on the *A* scale corresponds to 5 units on the *S* scale. Therefore .028 on the *A* scale corresponds to  $5 \times .028$ , or .14, on the *S* scale. Adding this amount to 32, we have 32.14 as the score value corresponding to 4.028 on the *A* scale. This is exactly the value we obtained by the first method, as, of course, it should be.

## ORDINARY SOLUTION

$$1 \times 52 = 52$$

$$5 \times 47 = 235$$

$$4 \times 42 = 168$$

$$5 \times 37 = 185$$

$$5 \times 32 = 160$$

$$8 \times 27 = 216$$

$$2 \times 22 = 44$$

$$5 \times 17 = 85$$

$$1 \times 12 = 12$$

---


$$36 \qquad 1157$$

$$1157 \div 36 = 32.14$$

(mean score)

## SOLUTION BY SUBSTITUTION

$$1 \times 8 = 8$$

$$5 \times 7 = 35$$

$$4 \times 6 = 24$$

$$5 \times 5 = 25$$

$$5 \times 4 = 20$$

$$8 \times 3 = 24$$

$$2 \times 2 = 4$$

$$5 \times 1 = 5$$

$$1 \times 0 = 0$$

---


$$36 \qquad 145$$

$$145 \div 36 = 4.028 \text{ (mean aux. val.)}$$

$$.028 \times 5 = .14$$

$$32 + .14 = 32.14 \text{ (mean score corresponding)}$$

We might have used the auxiliary values, 1, 2, 3, etc., to 9 instead of 0, 1, 2, etc., to 8. Or we might have used 4, 5, 6, etc., to 12, or 2, 4, 6, etc., to 18. No matter what scale we use, the mean value is bound to come out at the same *point* on the scale, although it may have a different numerical value, and we should always get 32.14 as the corresponding score value.

This method of finding a mean by substitution may seem to you now to be so complicated as to be impractical, but you will come to see that it is in many cases a great convenience.

**EXERCISE 15.** Make a rough sketch like Figure 9, placing the numbers 7, 12, 17, etc., up to 52 on the left, and represent these score values by auxiliary values 1, 2, 3, etc., up to 10. Find the mean of the auxiliary values corresponding to the scores of Grade 4B (Table 11), using mid-interval values in lieu of actual scores. Convert the mean auxiliary value into terms of scores as in the calculation above. First find the score corresponding to the whole number, then remember that 1 unit on the *A* scale equals 5 units on the *S* scale.

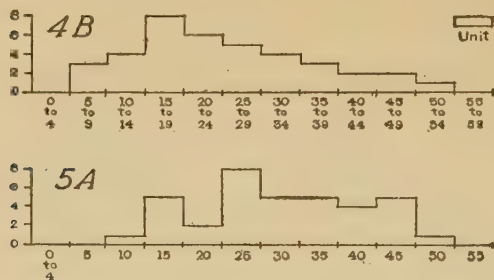


FIG. 10. Histograms of grouped distributions of scores of Grades 4B and 5A in the mental-ability test.

**Histograms.** A grouped distribution may be represented by a histogram in the same way as an ungrouped distribution, except that each rectangle represents the frequency of scores falling within a certain interval rather than the frequency of scores of one value.

Histograms representing the distributions of scores of Grades 4B and 5A are shown in Figure 10. You will see that in this case the unit of area is not a square but a rectangle one fourth as high as it is long, as shown by the small rectangle in the corner, marked *unit*. It is sufficient to indicate the beginning of each interval of score, except the first, as shown in the lower scale.

**EXERCISE 16.** Make an enlarged copy of Figure 10 on cross-section paper and add on the same sheet the histograms to represent the distributions of scores of Grades 5B and 6A given in Table 11. Let corresponding score intervals be aligned vertically so that the relative position of the histograms may be seen at a glance.

(For further practice see Appendix III, page 307.)

**The graph of a grouped distribution.** We are now ready to apply our methods of representing a distribution by a step graph and line graph (line of diagonals) to the case of a grouped distribution. The line graph is, of course, the simpler and more practical of the two, but it is well to under-

stand the principle of the step graph as an aid to the clearer appreciation of the line graph.

The relation between the histogram, step graph, and line graph is exactly the same in the case of an ungrouped distribution. Thus Figure 11 shows the step graph and histogram of Grade 4B (frequencies of scores in the mental-ability test from Table 11) in the same relative positions as in Figure 4, and the line of diagonals.

**Finding the median of a grouped distribution.** Let us now consider the method of finding the median score of Grade 5A in the mental-ability test. Since there are 36 cases as before, the median score is halfway between the 18th and 19th scores when arranged in order. If we begin at the left and count the marks indicating the 5A scores in Table 11, we find that the 18th and 19th scores are the second and third in the group of five in the compartment under 30-34.

We may now either make an estimate as to what the median score is, or we can look up those five papers and

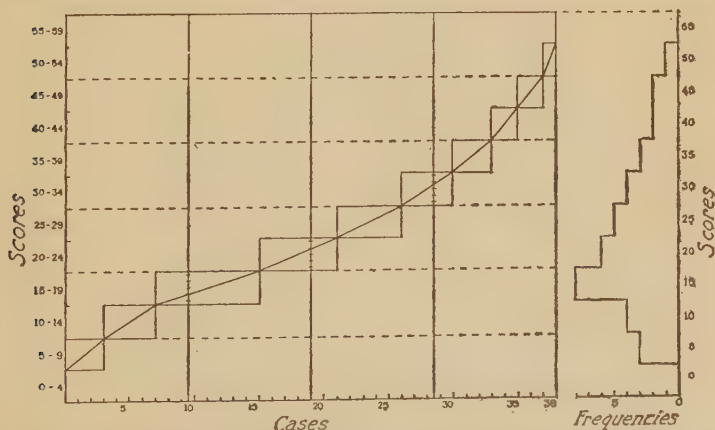


FIG. 11. The step graph, line of diagonals, and histogram representing the grouped distribution of scores of Grade 4B in the mental-ability test; illustrating the method of reading a line graph.

find the exact scores that are the second and third in order of size.

**Finding the median by estimate.** The method of estimating the median is to assume that the scores in the group, within which the median is found to be, are distributed within that group at approximately equal intervals. Thus, in this case, since there are five scores in the group containing the median, we may assume that these five scores are most probably 30, 31, 32, 33, and 34. Then, remembering that the median is halfway between the 2d and the 3d scores, counting upward, this means that the median is assumed to be halfway between the scores of 31 and 32. We should therefore say that the median score of Grade 5A is  $31\frac{1}{2}$ ; or, to avoid fractions,<sup>1</sup> we might merely say that it is 32. (When the range of scores is as great as in this case, a half point is hardly worth taking into account.)

**Finding the median by consulting original papers.** Now let us go back to the original scores and see how nearly the actual medians correspond to those we obtained by the method of estimate. If we look up the five actual scores in Grade 5A that fell into the interval 30–34, we find these to be 30, 30, 31, 32, and 34. The median is, therefore, halfway between 30 and 31, and is  $30\frac{1}{2}$ ; or, disregarding fractions, we should say that the median is 30. This happens to come out one point less than the median we got by estimate.

**Problem:** Find the median score of Grade 5B in the mental-ability test (1) by the method of estimate and (2) by the method of consulting the original papers (scores in Table 10).  
**Solution:** (1) Since there are 47 cases, the median score is the 24th score, counting from either end. Counting from the lower end, the 24th score comes out the third one of the four scores in the interval 35–39. Let us assume that these four

<sup>1</sup> When rounding off the fraction  $\frac{1}{2}$ , it is customary to take the nearest even number.

scores are most probably 35, 36, 38, and 39 — this is about as nearly equally as we can space them <sup>1</sup> — in which case we should say that the median is most probably 38. (2) Looking up the four actual 5B scores which fell into the interval 35–39, we find these to be 36, 37, 38, and 39. The third score in this case is 38, which corresponds exactly to our guess.

**How to find the median of a grouped distribution graphically.** We can estimate the median of a grouped distribution very easily from a line graph. Thus, suppose we wish to estimate the median of Grade 4B from the line graph in Figure 11. This is done as follows:

We draw a vertical line through the exact middle of the graph, just as we did in the case of ungrouped distributions. Since there are 38 cases, this line, of course, must pass between the 19th and 20th unit spaces along the horizontal scale.

This line is drawn in Figure 11 and passes through the rectangle representing the six scores that fell in the interval 20–24. We have now to estimate whether the median is 20, 21, 22, 23, or 24.

Let us take an enlarged view of the rectangle containing the median score. This is shown in Figure 12. In the large rectangle are drawn small ones representing the six separate

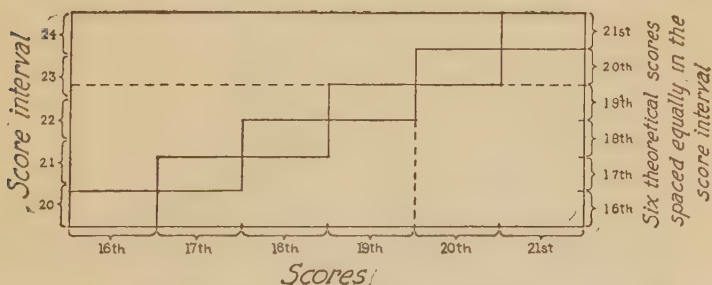


FIG. 12. An enlarged view of the rectangle in Figure 11 containing the median score; illustrating the method of finding the median.

<sup>1</sup> The basis for the choice of these four scores is given below.

scores that fell in the interval 20–24 just as these would be represented in the step graph of an ungrouped distribution, assuming them to be spaced equally in the interval. (This assumption is always made in dealing with grouped distributions.)

Of course six actual different scores cannot be spaced equally in an interval of only five scores; but we may divide the scale distance representing the score interval 20–24 into six equal parts, as shown at the right in Figure 12, and draw the six small rectangles to fit these six divisions, as shown.

Now the point in the scale between the 19th and the 20th scores according to our six divisions comes at a height within the unit on the scale at the left representing the score of 23. Therefore, our best estimate as to the value of the actual median is 23. Or we may say that the median score, found graphically, is 23.

It is not necessary, of course, to draw the small rectangles shown in Figure 12. Thus, Figure 13 shows how the same result might be obtained without the rectangles. All that is necessary is to note the point where the median line cuts the diagonal and compare the height of this point with the vertical scale. In Figure 13 the heights of the mid-points<sup>1</sup> of



FIG. 13. Showing the method of finding the median as in Figure 12 without small rectangles and illustrating the use of mid-points of the scale units.

<sup>1</sup> See page 50 for the reason for using mid-points.

the five scale units are indicated by cross lines on the median line and you will see that the height of the point where the diagonal crosses the line is nearest the height of the cross line representing the score of 23.

If there had been an odd number of pupils in Grade 4B and the median score had been the 19th instead of between the 19th and the 20th, then the vertical line in Figures 12 and 13 would have been drawn through the middle of the horizontal unit marked 19th, but the method of estimating the median would have been just the same. We should have found the point where the vertical line cut the diagonal and compared the height of this point with the vertical scale.

Referring now to Figure 11, you will see that the median vertical line (through the middle of the graph) has been graduated with cross lines representing the mid-points of the five units of the scale interval 20–24, the same as in Figure 13, and that the line of diagonals passes nearest to the cross line representing the score of 23, the same as in Figures 12 and 13. By this method the median score of the distribution represented by this line graph is found to be 23. We shall no longer think of ourselves as seeking to estimate the actual median but shall consider that we are *finding* the median by the graphic method. In other words, we shall use a theoretical median instead of the actual median in the case of grouped distributions.

**Finding quartile scores graphically.** The upper and lower quartile scores of the distribution of scores of Grade 4B are found by means of the line of diagonals in exactly the same way as the median. Thus, to find the quartile scores, we draw two more vertical lines which, with the middle line, will divide the width of the graph into four exactly equal parts. By graduating<sup>1</sup> these lines in the vicinity of the line

<sup>1</sup> In practice this graduating can be done very easily by means of charts which are provided.

of diagonals, we see that the lower quartile line cuts the line of diagonals at a point nearest the graduation representing a score of 16. So we should say that the lower quartile score of the distribution is 16. The upper quartile line cuts the line of diagonals at a point nearest the graduation representing a score of 33. So we should say that the upper quartile score of the distribution is 33.

**Calculating the median of a large group.** If you observe Figures 12 and 13 carefully, you will see that finding the median graphically in this case virtually consists in finding the point  $\frac{4}{6}$  of the way from  $19\frac{1}{2}$  to  $24\frac{1}{2}$ . (6 = number of cases in interval and 4 = number of cases below the median point.)<sup>1</sup> Let us suppose there had been 50 cases in the interval and the median had been the 37th counting upward, or had been halfway between the 37th and 38th scores. In either of these cases it is sufficiently accurate for all ordinary purposes to find the point  $\frac{37}{50}$  of the way between  $19\frac{1}{2}$  and  $24\frac{1}{2}$ , or, in other words, to find  $19\frac{1}{2} + \frac{37}{50}$  of 5. (5 = interval of score.)<sup>2</sup> In that case the median would be taken as  $19\frac{1}{2} + 3.7$  or 23.

The "median point" in the distribution of scores of a large group may be calculated by the formula,

$$(\text{Med. Pt.}) = \left( \begin{array}{c} \text{Lowest point} \\ \text{in interval} \end{array} \right) + \left( \frac{\text{No. of cases to median}}{\text{No. of cases in interval}} \times \text{interval} \right),$$

in which the "lowest point in interval" means a point  $\frac{1}{2}$  unit below the lowest score of the interval;<sup>3</sup> "No. of cases to median" means the number of cases in the interval to and including the median score (if an odd number) or the number of cases in the interval below the median point (if an

<sup>1</sup> If the 19th case had been the median, the graphic solution would amount to taking  $3\frac{1}{2}$  sixths of the interval.

<sup>2</sup> To be strictly accurate in the first case you should take only  $36\frac{1}{2}/50$  of 5 to add to  $19\frac{1}{2}$ .

<sup>3</sup> The reason for taking a point  $\frac{1}{2}$  unit below the lowest score is discussed under the heading, "A point of theoretical interest."

even number), and “interval” means the number of units of score in the interval. For the median score, take the whole number nearest the median point.

*Problem:* Suppose you encountered a group so large that there were, say, 758 cases in the interval containing the median. Suppose the median was the 236th case in the interval counting upward and that the interval containing the median was the interval 160–169. What would the median score be? *Solution:*

$$\begin{aligned}\text{Median point}^1 &= 159.5 + \frac{236}{758} \times 10 \\ &= 159.5 + 3.1 \\ &= 162.6\end{aligned}$$

The median score is therefore 163, the nearest whole number.

**More precise statements of the median.** When the median (or other percentile) scores of two distributions differ by but very few points or possibly less than one unit on the vertical scale, it is necessary to state the median more precisely than merely in terms of whole numbers. Thus, in Figure 11, for example (see Figure 13 also), since the diagonal cuts the vertical line at a point about .9 of the distance from graduation 22 to graduation 23, we may say, more precisely, that the median of the distribution is 22.9. Similarly the upper and lower quartile scores in Figure 11 may be considered as 32.7 and 16.0, respectively.

**A point of theoretical interest.** A point that is likely to arise in the drawing and reading of line graphs is this: In finding the median in Figure 11, for example, why is the *point* representing a score of 20 considered as being the mid-point of the vertical scale unit representing the score of 20, etc. (see also Figure 13), instead of the lower limit of this scale unit?

<sup>1</sup> To be strictly accurate, we should take  $235\frac{1}{2}/758$ , but you can see that the difference would be entirely negligible.

What has been done is merely so to graduate the vertical scale that when the median point falls within the unit of the interval representing the score of 20, it will be *nearest* to the graduation representing that score, and the same for the other units. In one sense, therefore, a score of 20 is thought of as extending from  $19\frac{1}{2}$  to  $20\frac{1}{2}$ .

**Some theoretical considerations.** In most books on statistical method the subject of the median is treated as of considerable importance, and the writer believes it will be worth while to discuss some theoretical points that seem not to be well understood. The reader not interested may skip this discussion.

It is contended by some that when finding the median of a grouped distribution of scores <sup>1</sup> the interval 20 to 24, for example, should be considered as extending *not* from 19.5 to 24.5 as we have assumed, but from 20.000 to 24.999 + (virtually from 20 to 25), on the assumption that a score of 20 means 20 problems solved and another one probably partly solved — anywhere from only just begun to .999 done, say, and the same for 21, 22, 23, and 24.

Now it is obvious that a pupil making a score of 20 has very probably partly solved a 21st problem and that he was most probably about halfway finished when time was called, let us say; so that in terms of actual amount of calculation or mental activity in a test, a score of 20 represents in reality most probably  $20\frac{1}{2}$  problem units. Of course we could add  $\frac{1}{2}$  to every score if we wished, in order that all our scores would be expressed in *problem units of mental activity*, instead of in terms of *problems solved*, but obviously that would be entirely impractical and needless. We can always interpret a score of 20 as meaning  $20\frac{1}{2}$  problem units of mental activity, if we wish.

Now those who contend that we should consider the inter-

<sup>1</sup> Therefore the hypothetical or estimated median.

val 20–24 as extending from 20.000 to 24.999 also say that if five scores fall in this interval and the median score is the middle one of the five, the median value is the mid-point of the scale distance 20.000 to 24.999, or 22.5. Obviously, this is in terms of what we have called problem units of mental activity and not problems solved.

But let us suppose that a school principal has made a report saying that the median of a certain class is  $22\frac{1}{2}$  (found as above), and let us suppose that Johnny Jones in the class has made a score of 22. Was his score equal to the median? Apparently not, according to the figures; but in reality yes, exactly. The median is in problem units of mental activity, and Johnny's score is in problems solved. If it were translated into problem units of mental activity, it would be  $22\frac{1}{2}$  too.

Do you now see why it is misleading to express medians in terms of problem units of mental activity and scores in terms of problems solved? What we need is a method of calculating the median that gives the median in the same terms as those in which the scores themselves that are to be compared with the median are expressed. That is why it is necessary to consider a score of 20, for example, as if it had a range from 19.5 to 20.5 instead of from 20 to 20.999. If five scores fall in the interval 20–24 and the median is the middle one, it will be found to be, by this method, 22. And that is exactly what the middle score of the five, 20, 21, 22, 23, and 24, is. The uncompleted fractional part of a problem is always taken for granted anyway. It does not need to be expressed.

### QUESTIONS

1. If you had given a test and found the scores of your pupils to extend from 86 to 243, what sized interval would you choose to group the scores for the purpose of distributing them?
2. Can you find the average of 162, 165, 160, 161, and 167 mentally?

## 52 *Statistical Method in Educational Measurement*

3. As the size of a group is increased, do you think the median found graphically more nearly or less nearly approximates the actual median?

4. Suppose the median of a certain distribution has been calculated, assuming the interval containing the median be extended from 50 to 59.999 and that the median is stated as being 56.3. Suppose that Frank has made a score of 56 in the test. Is his score below the median?

## CHAPTER FIVE

### THE PERCENTILE GRAPH

THERE is still a certain lack of convenience in both the step graph and the line of diagonals of Figure 11 in that in order to find the median and quartile <sup>1</sup> scores or the various percentile scores it is necessary to divide the width of the graph into quarters or tenths or hundredths.

**The percentile graph.** In order to overcome this inconvenience we may draw the rectangles or line of diagonals on a chart in which the horizontal distance that the rectangles or line will occupy is already graduated into 100 units. Such a chart is called a percentile chart or *percentile graph*.<sup>2</sup> A percentile graph designed especially for use with the Otis Self-Administering Tests of Mental Ability is included in each package of these tests. A reproduction of the percentile graph is shown in Figure 14.<sup>3</sup> The step graph and line of diagonals of Figure 11 are shown drawn upon it. The method of drawing these is explained a little farther on. By comparing Figure 14 with Figure 11 we can see at a glance the greater convenience of the percentile graph.

**Some advantages of the percentile graph.** First of all, the horizontal distance occupied by the graph is already divided into quarters for finding the median and the upper and lower quartiles. Moreover, this distance is graduated above and below into units of one per cent of the distance and the 10-

<sup>1</sup> See page 31 for meaning of quartile scores.

<sup>2</sup> It might be more correct to refer to the chart as a "percentile chart" so long as nothing has been drawn upon it, and to call it a percentile graph when it contains a line of diagonals or other representation of a distribution. The term "percentile graph," however, has come to apply to the blank chart as well.

<sup>3</sup> For the idea of providing an especially prepared chart for drawing line graphs the writer is indebted to Professor W. S. Miller, of the University of Minnesota, author of the Miller Mental Ability Test. The percentile graph furnished by Professor Miller with his test served as the point of departure in the construction of the percentile graphs described in this book.

percentile, 20-percentile, etc., lines are drawn so that scores corresponding to these percentiles can be easily found.

The 50-percentile line is already graduated to show the mid-points of the units of the vertical scale representing the various scores. By means of these graduations the median score is instantly seen to be 23. The dotted lines show  $Q_3$  and  $Q_1$  to be respectively 33 and 16 as before. A further advantage of the percentile graph is that by it all lines of diagonals are brought to the same width. This admits of lines of diagonals representing two or more distributions being drawn on the same graph. Other advantages of percentile graphs are brought out later.

The general utility of the percentile graph. If you have made a casual perusal of this book, you have doubtless noted that the discussion very frequently involves the use and interpretation of lines and curves drawn upon the percentile graph. The fact is that the percentile graph is perhaps the most useful and most convenient single chart that may be employed in the interpretation of scores and in research in the field of mental and educational measurements. If you do not master these early chapters explaining the percentile graph, you may never fully appreciate its value. On the other hand, if you gain a thorough understanding of the percentile graph you will have at your command a means of interpreting scores and making statistical comparisons that is far simpler, quicker, and easier than many of the methods now in use. It is, indeed, something comparatively new,<sup>1</sup> but its usefulness is becoming more and more appreciated. You will be amply repaid for all the time you spend in becoming thoroughly acquainted with it.

<sup>1</sup> The idea of drawing percentile curves is not new. These are described and illustrated by Yule (G. Udny Yule, *An Introduction to the Theory of Statistics*, pages 151-153; J. B. Lippincott Company) under the name of *ogives*. The development of their use, however, has awaited the appearance of an especially prepared chart which would reduce the drawing of percentile curves to a simple and easy operation.

## PERCENTILE GRAPH For Intermediate and Higher Examinations

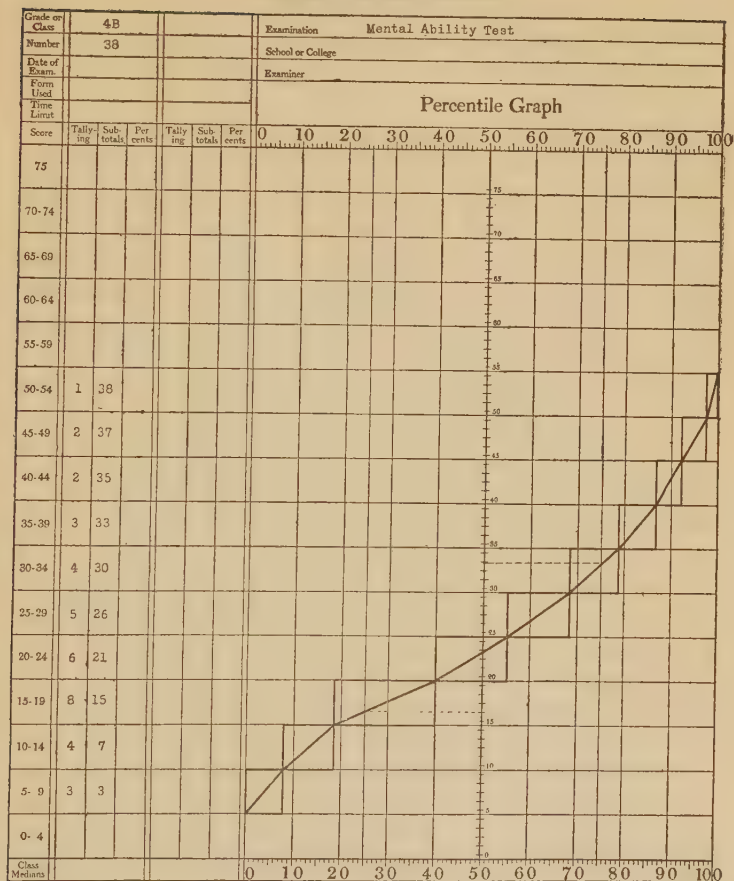


FIG. 14. The step graph and line of diagonals of Figure 11 drawn on a percentile graph.

How to draw a line of diagonals on a percentile graph. The percentile graph for the Otis Self-Administering Tests of Mental Ability is designed so that the scores may be dis-

tributed directly upon it. For this purpose the score intervals, 0-4, 5-9, etc., are printed in a column at the left. The method of drawing a line of diagonals on the percentile graph will be shown with reference to the distribution of scores of Grade 5B.

**Distributing the scores.** Provision is made for distributing the scores of two groups of individuals on one percentile-graph sheet, and from these distributions two lines of diagonals may be drawn. This does not mean, however, that only two lines may be drawn on one graph. The scores of additional groups may be distributed on other percentile-graph sheets or on any sheet of paper and as many lines drawn on one graph as may be conveniently distinguished.

In one of the columns headed "Tallying" distribute the scores of a class by putting a short mark opposite the interval of score within which the score of each individual falls. The percentile graph in Figure 15 shows that in Grade 5B five individuals had scores between 15 and 19, five had scores between 20 and 24, four had scores between 25 and 29, etc. If the distributing has been done already in a table such as Table 11, then of course figures representing the frequency of scores in the various intervals can be entered directly in the column under Tallying.

**Steps toward locating the points in the graph.** In locating the points in the graph to mark the ends of the diagonals we must use a somewhat different method from the one we have used so far, since the horizontal units representing the cases must be such that the whole number will occupy just the width of the graph. In other words, if there are 38 cases as in Grade 4B, the width of the graph must be divided into exactly 38 equal parts. Each  $\frac{1}{38}$  of the width of the graph will then represent one case. The rectangles in Figure 14 are based on a unit  $\frac{1}{38}$  of the width of the graph.

In the case of Grade 5B there are 47 cases; therefore we

## PERCENTILE GRAPH For Intermediate and Higher Examinations

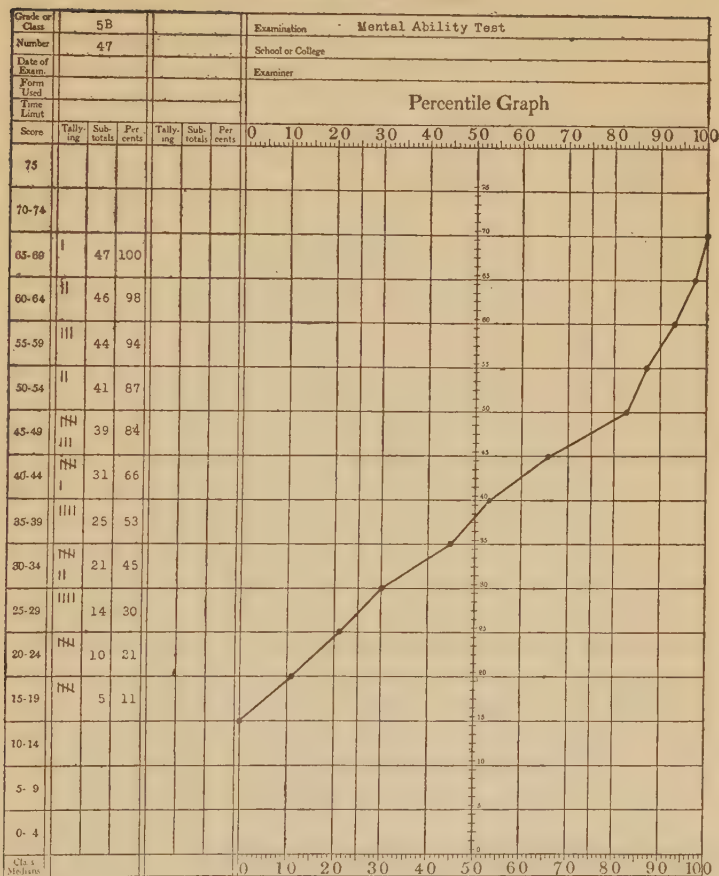


FIG. 15. Illustrating the method of drawing a line of diagonals on a percentile graph.

must contrive to divide the width of the graph into 47 equal parts, one unit to represent each of the 47 cases. (See Figure 15.) There are two ways to do this. One is to convert

the various 47ths into per cents and use the scale from 0 to 100, and the other way is to use a scale in which the width of the graph is divided into 47 equal parts. Both methods are described.

However we make this division, the first point, marking the beginning of the first diagonal, must be put, of course, on the left edge of the graph and on the horizontal line marking the lower limit of the lowest interval containing a score. In this case (Fig. 15) it is the interval 15–19. The second point, marking the end of the first diagonal, must be put in the next line above and at a distance to the right in this case *five* 47ths of the width of the graph. The third point must be put on the next line above and a distance of *five* 47ths to the right of the second point. To locate the fourth point we go to the next line above and over *four* 47ths, next up one line and over *seven* 47ths, etc., according to the frequencies.

**Cumulative errors.** If we located these points by measuring in each case from the point last placed, the positions of the points would be subject to *cumulative errors*. That is, if any one point were wrongly placed, all succeeding points would be affected by that error, and if errors were made in two or more measurements these would *accumulate* to affect all succeeding points. For this reason it is better to locate each point by direct measurement from the left edge of the graph rather than to locate each by measuring from the one just located. Thus, it is better to place the third point by measuring ten 47ths to the right of the edge, the fourth point by measuring fourteen 47ths to the right of the edge, etc. This necessitates finding these subtotals, 10, 14, 21, etc., of the frequencies.

**Finding the subtotals.** To find these subtotals we have but to begin at the bottom of the column of frequencies and place in the square at the right of each frequency the sum of the frequencies up to and including those in that group. In

Figure 15 the lowest frequency is 5. The sum (subtotal) of the first and second is 10, the sum of the first, second, and third is 14, etc., up to the last frequency, where the sum (47), of course, should equal the number of pupils in the grade, as entered at the top of the column.

**Converting subtotals to per cents.** As suggested above, one of the ways to plot the points marking the ends of the diagonals in a line graph, when there are 47 cases, for example, is to change 47ths to 100ths and plot by means of the percentile scales. Thus, as shown in the column headed Per Cents in Figure 15, five 47ths is 11 per cent, ten 47ths is 21 per cent, fourteen 47ths is 30 per cent, etc., and forty-seven 47ths, of course, is 100 per cent.

**Plotting the points by the percentile scale.** The second point in Figure 15 is to be placed 11 units to the right of the left-hand margin of the graph, the third point 21 units to the right of the margin, the fourth point 30 units to the right of the margin, etc.; the last, being 100 units to the right of the left margin, will, of course, fall in the right-hand margin.

**A table of per cents.** For convenience in employing the method that makes use of per cents when the number of cases is not over 60, a Percentage Table (Table I in Appendix II) is given, by means of which any subtotal may be converted into a per cent by consulting the proper column. Thus, by consulting the column headed 47ths it will be seen that five 47ths is 11 per cent, ten 47ths is 21 per cent, etc.

This table may be used, of course, for other purposes than merely in connection with a percentile graph. Whenever it is desired to divide one number under 60 by another, the table may be used. Thus, dividing 37 by 56 is the same as finding what per cent 37 is of 56, or, in other words, converting  $\frac{37}{56}$  to a per cent. So we find the number in the 56th column opposite 37 at the left. This is 66; therefore 37 divided by 56 = .66.

**Plotting points by means of the scale chart.** As suggested above, there is a second method of plotting the points in a percentile graph marking the ends of the diagonals in a line graph — a method making use of a scale in which the width of the graph is divided into as many equal parts as there are cases.

For this purpose a scale chart, as shown at *A* in Figure 16, accompanies each percentile graph. To plot the points for Grade 5B, for example, as in Figure 15, simply lay a strip of paper along Scale 47 (shown at *X* in Figure 16) and mark the distances from an initial point on the paper, equal to 5 units of Scale 47, 10 units, 14 units, 21 units, etc., according to the subtotals. Then lay the strip of paper on the percentile graph and lay off the successive distances on the proper cross lines.

If a separate scale chart is at hand, this may be folded on the proper scale and applied directly to the percentile graph for measuring off the distances.

It will be seen that there are scales in Scale Chart A having 40 units, 41 units, etc., up to 100 units. In a case such as that of Grade 4B, for example, where there are only 38 cases, no one of these scales can be used directly. However, if we use Scale 76 we may count two units as one and thus have the width of the graph divided into 38 equal parts. Similarly, if we had more than 100 cases, say 138, we could use Scale 69, letting each unit represent two cases, or, in other words, two 138ths of the width of the graph. If we had 550 cases, we could use Scale 55 and let each unit represent ten cases, etc. In these ways the scale chart may be used for any number of cases.

**Drawing the line of diagonals.** After the points have been plotted by either of the above methods, there remains merely to connect these by lines, as shown in Figure 15, to complete the line of diagonals which represents the distribu-

## SCALE CHARTS

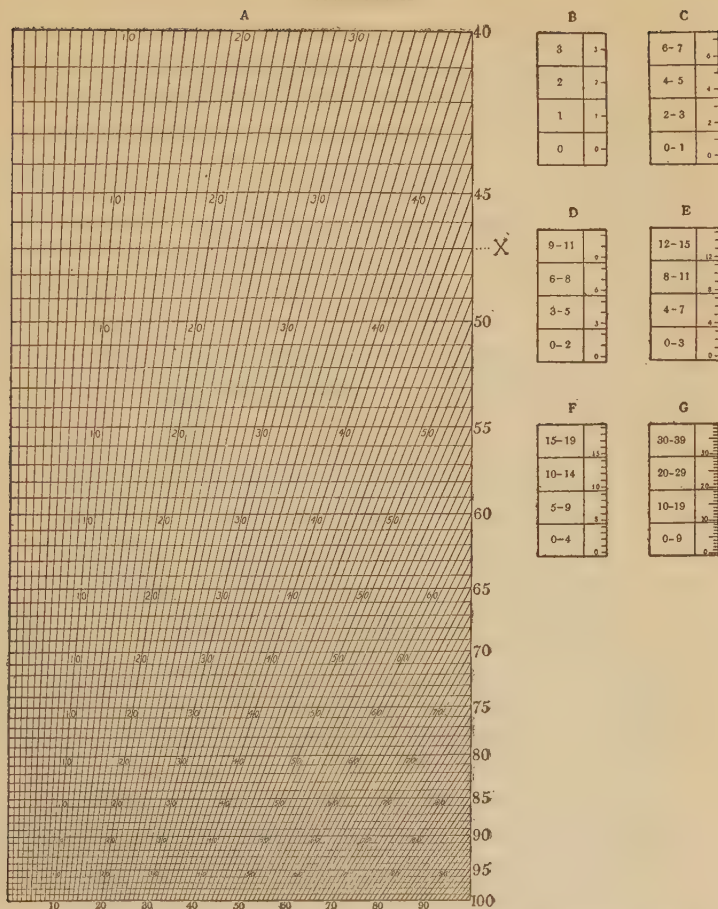


FIG. 16. The scale charts furnished with the Universal Percentile Graph.  
(Actual size, 8 by 10 inches.)

tion. By means of this line of diagonals the median, quartiles, and any percentile scores of the distribution may be found by use of the percentile lines and scales.

**The Universal Percentile Graph.** In order to provide a single percentile graph suitable for all tests — one that might be used with any intervals of score — the Universal Percentile Graph has been devised.<sup>1</sup> This differs from the percentile graph for the Otis Self-Administering Tests of Mental Ability in only a few minor details. First of all, provision is made for distributions having twenty-one intervals of score instead of but fifteen. (See Figure 17.) Then, in order that distributions having various intervals of score may be represented on it, the spaces for intervals of score are left blank and are to be filled in by hand.<sup>2</sup> Thus in some cases it might be convenient to use the intervals 0-2, 3-5, 6-8, etc., and in others it might be desirable to use 0-9, 10-19, 20-29, etc.

**Scale Charts B, C, D, E, F, and G.** For the same reason, the 50-percentile line has not been graduated. In order that this line may be graduated in any desired units, auxiliary scale charts as shown in Figure 16 are provided with each Universal Percentile Graph. By folding the appropriate one of these charts on the right-hand edge and applying it to the 50-percentile line or any other percentile line, this line may be graduated to show the mid-points of the spaces representing the units composing the interval. For one unit per interval use Scale Chart B, for two units per interval use Scale Chart C, etc.

**Drawing two or more lines on one graph.** Figure 17 shows the lines of diagonals representing the distributions of scores of both Grades 4B and 5B. It is not possible to draw these on the same graph in this way without bringing them to the same width, and the special merit of the percentile graph is that it accomplishes this. Occasionally there is some inter-

<sup>1</sup> Published by World Book Company, Yonkers-on-Hudson, New York. Sold in packages of 25, each graph complete with scale charts and directions for use.

<sup>2</sup> The intervals of score in Figure 17 were entered with a typewriter.

UNIVERSAL PERCENTILE GRAPH

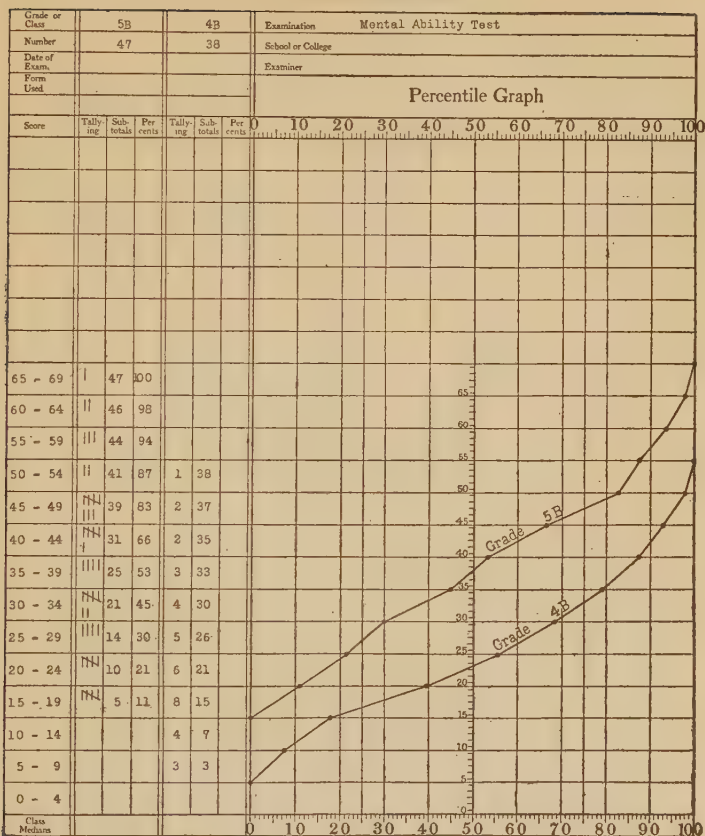


FIG. 17. The Universal Percentile Graph. (Actual size, 8 by 9½ inches.)

ference of lines, but this does not often happen and causes no serious inconvenience. There are, on the other hand, distinct advantages in representing two or more distributions on the same graph. These are discussed in a later section.

**EXERCISE 17.** Procure a copy of the Universal Percentile Graph and draw lines of diagonals to represent the distribution of scores of Grades 5A and 6A in the mental-ability test. Distribute the scores of Grade 5A directly from Table 10, but merely write the frequencies of Grade 6A from Table 11. Use the Percentage Table (Table I) in the case of Grade 5A and the scale chart (A) in the case of Grade 6A.

**EXERCISE 18.** Find the median score, upper and lower quartile scores (i.e., the 75- and 25-percentile scores), 10- and 90-percentile scores of the distributions of Grades 5A and 6A found in Exercise 17.

(For further practice see Appendix III, page 307.)

The drawing of a line of diagonals by means of the percentile graph is subject to refinement of a practical nature. The method is described in Chapter VII. We shall do well to consider first the effect of grouping upon a distribution, the general shape of lines of diagonals and histograms, the law of normal distribution, and the nature of a sample.

**The effect of grouping.** We have already made some investigation of the relation between an actual median and a hypothetical median found from a line graph. Let us investigate further the effect of grouping on a distribution.

In Figure 18 a distribution of scores of 140 pupils of Grade 5 (5A and 5B) in the mental-ability test is represented at *A* by (1) a step graph in which the frequency of each score is represented by a separate small rectangle and (2) by a step graph in which the scores are grouped in intervals of five and the frequencies are represented by large rectangles. Similarly, at *B* the same distribution is represented (1) by a full-line histogram showing the frequency of each score and (2) by a dotted-line histogram showing only frequencies of score in intervals of five units. The rectangles composing the full-line histogram are drawn to a scale five times that of the rectangles in the dotted histogram. (See the scales at the top and bottom.)

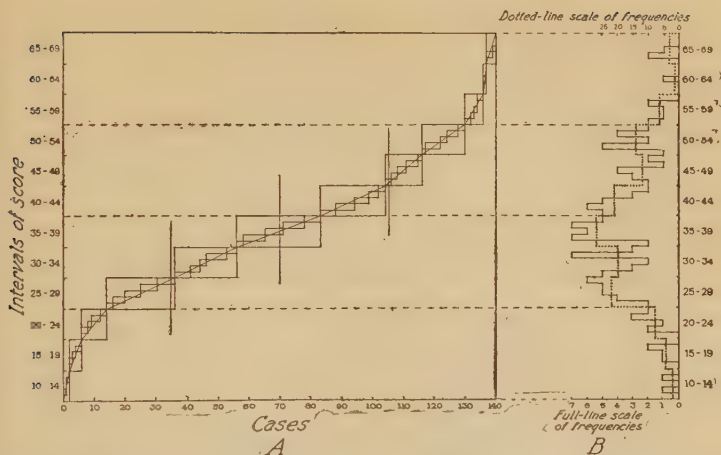


FIG. 18. Histograms and step graphs of grouped and ungrouped distributions of scores of 140 pupils of Grade 5A in the mental-ability test, with line of diagonals of step graph of grouped distribution; showing the smoothing effect of grouping.

The diagonals of the large rectangles at *A* are also drawn, and it may be plainly seen how closely this line of diagonals comes to passing through every small rectangle. The median line, for example, cuts the line of diagonals at a point within one of the small rectangles, and the upper and lower quartile lines also cut the line of diagonals at points within the small rectangles. In other words, so far as the finding of the actual values of  $Q_3$ ,  $M$ , and  $Q_1$  are concerned, nothing is lost in this case by grouping.

**The smoothing effect of grouping.** The dotted-line histogram at *B* shows the smoothing effect that grouping has upon the distribution. For example, whereas actually the frequencies of scores of 30, 31, 32, 33, and 34 are, respectively, 5, 3, 2, 7, and 3, the dotted-line histogram virtually assumes that the frequency of each of the five scores is 4 — the average of the five frequencies. Grouping, in other words, vir-

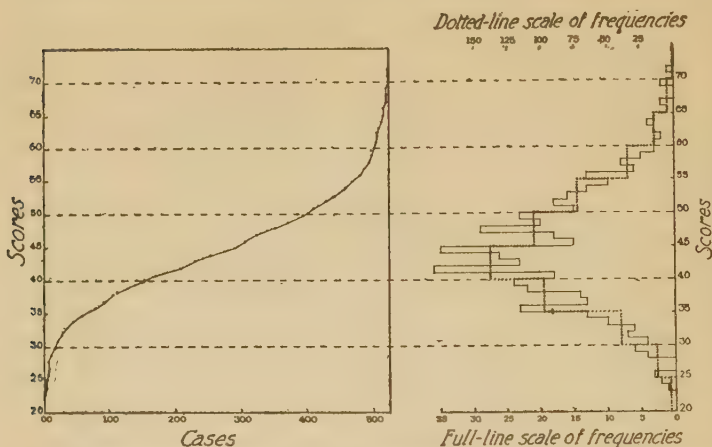


FIG. 19. Showing the tendency for the line of diagonals of a large distribution to approach a smooth curve, and the smoothing effect of grouping upon a histogram.

tually averages the frequencies of scores grouped in each interval of score.

Let us consider the line of diagonals and the histogram of a larger distribution. Figure 19 shows the line of diagonals and the histogram (with magnified rectangles) representing a distribution of 523 scores of college freshmen.<sup>1</sup> Here we see a marked tendency for the line of diagonals to resemble a smooth curve. The smoothing effect of grouping is also shown by the dotted-line histogram.

**The general shape of a graph.** You have doubtless noted that a step graph or line of diagonals tends to take a characteristic shape. It tends to be most nearly level in the middle and to curve rather sharply up and down at the ends, approaching the vertical.

<sup>1</sup> Scores in the Higher Examination of the Otis Self-Administering Tests of Mental Ability, using a 20-minute time limit.

The writer is indebted to Professor Grover Hooker of the University of Colorado for these scores.

A histogram, on the other hand, tends to be tallest (or thickest) in the middle and to tail off at the ends. The next chapter discusses the characteristic shape of histograms and lines of diagonals, and throws light on a number of theoretical and practical considerations bearing upon the drawing of graphs of distributions.

**Finding a percentage graphically.** Note that Scale Chart A can be used also for finding percentages or, in other words, for dividing one number by another. For example, suppose that we wish to find what per cent 72 is of 96 or, in other words, to divide 72 by 96. If we folded Scale Chart A on scale 96 and applied it to the scale at the top or bottom of the Percentile Graph, we should find the point 72 on scale 96 to be directly opposite 75 on the Percentile Graph. This shows that 72 is 75 per cent of 96 or, in other words, 72 divided by 96 = .75.

Similarly, to find, say, 68 per cent of 75, we have but to apply the Percentile Graph scale to the scale 75 on Scale Chart A and find the point on that scale opposite 68 on the Percentile Graph scale. This number is 51. Therefore, 68 per cent of 75 is 51.

### QUESTIONS

1. Which seems more convenient to you in drawing a line graph, to convert the subtotals into percentages or to use Scale Chart A?
2. Is one of these methods more accurate than the other?
3. Why do you think a line graph of an ordinary distribution of scores tends to be more nearly horizontal near the middle?

## CHAPTER SIX

### THE LAW OF NORMAL DISTRIBUTION

YOU have no doubt often heard the term “law of normal distribution” and may have seen frequently a bell-shaped curve which is variously referred to as a normal-distribution curve, normal-frequency curve, normal surface of distribution, normal surface of frequency, normal probability curve, etc. It is the purpose of this chapter to make clear the meaning of this curve and of the expression “law of normal distribution.”

In Figure 20 is shown a histogram representing a grouped distribution of the scores of 628 soldiers in the Army group intelligence test Alpha.<sup>1</sup> It will be seen that this histogram is somewhat more regular than the smoothed histogram in Figure 19. It is also more symmetrical. In fact, one can see a distinct tendency for each frequency on one side of the middle one to be balanced by a frequency of almost the same amount on the other side.

In Figure 21 is shown a histogram representing the scores of a group of 25,200 soldiers in the Army group intelligence

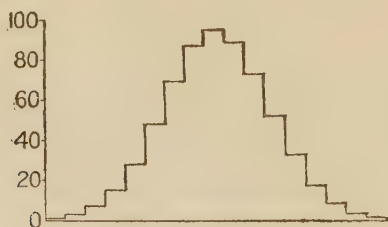


FIG. 20. A histogram representing distribution of the scores of 628 soldiers in the Army group intelligence test Alpha. (Note its symmetrical appearance.)

<sup>1</sup> Taken from an article by Percival M. Symonds in the February, 1923, number of the *Journal of Educational Psychology*, page 72.

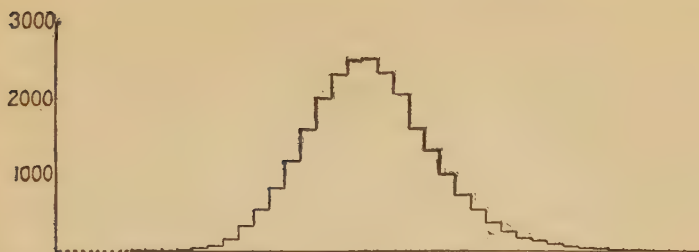


FIG. 21. A histogram representing the distribution of scores of 25,200 soldiers in the Army group intelligence test Alpha.

test Alpha. It will be seen that the histograms in Figures 20 and 21 are very similar in shape — one might almost say they are identical in shape; that is, they are both very nearly symmetrical, with frequencies of approximately equal amounts balancing each other on the two sides of the middle. In each histogram the difference between the amounts of two adjacent frequencies is greatest about halfway down either side, and in each case the difference between adjacent frequencies becomes less and less as we approach the base line on either side, or as we approach the middle of the histogram.

There is nothing accidental about the great similarity between these two histograms. If we should test the same number of individuals represented in the histogram in Figure 21 by any other intelligence test, such as the National Intelligence Test or the Herring Revision of the Binet-Simon Tests or the Terman Group Test of Mental Ability, or indeed by any achievement test such as the Stanford Achievement Test, we should almost certainly find that the histogram representing the scores of these individuals assumed a shape very similar indeed to that of the histogram in Figure 21. Even though there should not be as many intervals of scores, the same tendencies that have been pointed out would be manifest.

The law of probability in the flipping of coins. The law of normal distribution is the law governing a distribution that is the result of pure chance. Thus, let us suppose that we flip two coins and count the number of "heads" appearing. The number may be, of course, either two, one, or none. The possible ways the coins could fall are shown below :

	FIRST COIN	SECOND COIN	NO. OF HEADS APPEARING
First possibility	heads	heads	2
Second "	heads	tails	1
Third "	tails	heads	1
Fourth "	tails	tails	0

One of these possibilities results in the appearance of two heads, two result in the appearance of one head, and one results in the appearance of no heads. Now since there is no more reason for any one of these possibilities to occur than any other, they tend to occur in equal numbers. That is, out of 100 throws, the first possibility tends to occur 25 times, the second 25 times, and the same for the third and fourth. This means, of course, that 2 heads tend to appear 25 times, 1 head 50 times, and 0 heads 25 times.

No matter how many throws of two coins were made, the ratio of the throws of 2 heads, 1 head, and 0 heads would tend to be 1 : 2 : 1.

If three coins were flipped, the possibilities would be :

	FIRST COIN	SECOND COIN	THIRD COIN	NO. OF HEADS APPEARING
First possibility	heads	heads	heads	3
Second "	heads	heads	tails	2
Third "	heads	tails	heads	2
Fourth "	heads	tails	tails	1
Fifth "	tails	heads	heads	2
Sixth "	tails	heads	tails	1
Seventh "	tails	tails	heads	1
Eighth "	tails	tails	tails	0

The ratio of throws of 3, 2, 1, and 0 heads would tend in this case to be  $1:3:3:1$ .

Similarly, it may be shown that if there were four coins the ratio of throws of 4, 3, 2, 1, and 0 heads would tend to be  $1:4:6:4:1$ .

If we went on in this way up to ten coins, we should find the probability of the ratio of throws of 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, and 0 heads to be

$$1:10:45:120:210:252:210:120:45:10:1.^1$$

Figure 22 shows a histogram representing the frequencies of the various numbers of heads falling according to the law of probability when 24 coins are flipped. Note how closely the shape of this histogram resembles that of the histogram of actual scores in Figure 20.

The histogram in Figure 20 does not extend so far out at the ends as the theoretical one; but that is partly because the scale of possible scores did not extend that far and partly because the area of the theoretical histogram outside the limits of the histogram of Figure 20 is less than  $\frac{1}{1000}$  of the whole area, and since there were only 628 cases,  $\frac{1}{1000}$  of the number would be less than one score.

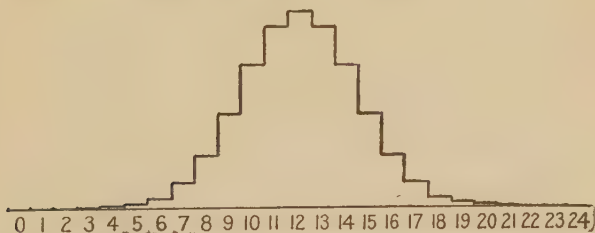


FIG. 22. A histogram representing the distribution of the numbers of heads falling according to the law of probability when 24 coins are flipped.

<sup>1</sup> These values are the same as the coefficients of the various terms in the expansion of the binomial  $(a + b)^{10}$ . The same is true for any other number of coins. The exponent of the binomial is the number of coins flipped.

**The chance element in scores.** Now if we think of the chance of a pupil knowing or not knowing the answer to a question in a test as something like the chance of a coin falling heads or tails, and hence the chance of a pupil making a certain score as something like the chance of there being a certain number of heads among 10 coins that have been flipped, we can see a reason for the distribution of scores of a homogeneous group of pupils resembling the distribution of numbers of heads among flipped coins.

We know, in fact, that there are hundreds of causes or influences operating to determine the score of a child in even the simplest test. A child's score is affected by all the days of instruction he has had, by all the factors making up his innate capacity to learn, by his interest in the subject as governed by the many ways in which it has or has not appealed to him as satisfying his needs, by the many factors making up the quality of his instruction, and, finally, — to some extent at least, — by the elements making up his physical condition and mood at the time of the test. All these factors, elements, and influences are of a nature such that the presence of any one of them in the case of one of the pupils of a group and its absence in the case of another is largely a matter of chance. So that, after all, there appears to be a marked similarity between the nature of the causes governing the distribution of scores of a group of pupils in a test and the causes governing the distribution of numbers of heads in a set of coins that have been repeatedly flipped. This reasoning, coupled with the fact that there is so close a correspondence in shape between the histograms of distributions of scores of homogeneous groups of pupils and the theoretical histograms of distributions of numbers of heads among flipped coins, as shown by Figures 20 and 22, leads us to say that such distributions of scores tend to follow the law of probability or *the law of normal distribution*.

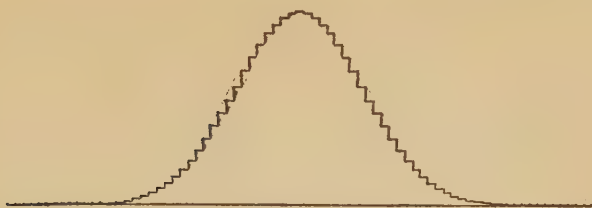


FIG. 23. A histogram representing the distribution of numbers of heads tending to appear according to the law of probability when a large number of coins are flipped many times.



FIG. 24. A theoretical normal surface of frequency showing the limit approached by a histogram representing a normal distribution when the number of cases is increased and the units made very small.

**The normal surface of distribution.** If we flipped a large number of coins many times, the histogram representing the numbers of heads appearing would tend to take the form shown in Figure 23. The histogram, in other words, would approximate a smooth, symmetrical curve such as that shown in Figure 24.

The curve in Figure 24, together with the base line, bound a surface which is properly called *the normal surface of distribution* or normal surface of frequency and may be thought of as a histogram representing the distribution of an infinite number of scores distributed exactly in accord with the law of normal distribution and so finely divided that no zigzag appearance is appreciable. Or it might be thought of as a smooth curve drawn through the mid-points of the horizontal

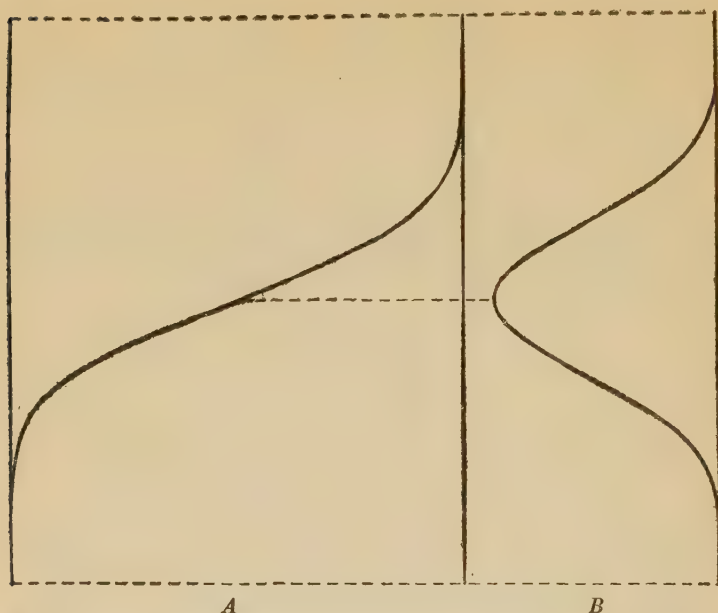


FIG. 25. Showing the correspondence between the normal surface of distribution (smooth histogram) and the normal curve of distribution (smooth line of diagonals).

lines of a histogram such as those in Figures 22 and 23, representing an infinite number of scores.

**The normal curve of distribution.** At *B* in Figure 25 is shown a normal surface of distribution similar to that shown in Figure 24, and at *A* is shown the theoretical normal line of diagonals corresponding to this surface. The smooth appearance of this line of diagonals is due to the fact that it also represents an infinite number of scores so finely divided that the length of each diagonal is inappreciable. Note its general shape, which is characteristic of the line graphs we have obtained heretofore.

The shape of the curve in Figure 24 is in conformity with a

definite mathematical law, or formula, and can be drawn to any degree of precision. This formula has been developed in accordance with the theory of probability, and for that reason the surface is sometimes known as the *probability surface*. Readers interested in the formula and mathematical properties of this curve are referred to more technical books, such as *Statistical Method*, by Truman L. Kelley.<sup>1</sup>

The histogram representing the distribution of scores in any test tends to be normal only when the scores are free from certain extraneous influences and only when equal increments of score are of approximately equal value throughout the scale.

**Skewed distributions.** There are, of course, many instances in which the free working of the law of normal distribution is disturbed by circumstances. For example, in the case of a test in which the scores may run from 0 to 75, if the median of a class tends to be 65, it is obvious that the highest score can be only 10 points above the median while the lowest score could be 65 points below the median. In such a

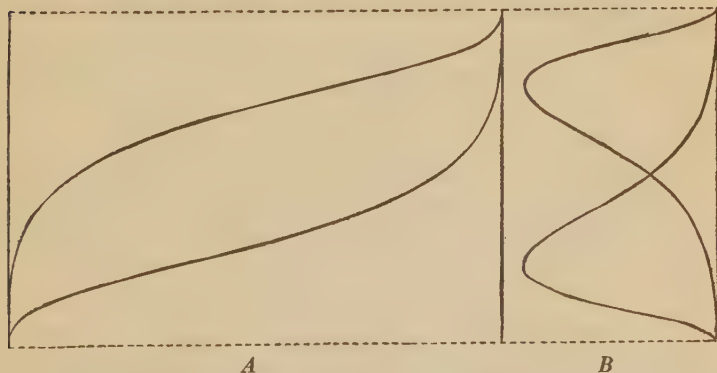


FIG. 26. Showing the correspondence between the surface of distribution (at B) and the curve of distribution (at A) of two skewed distributions.

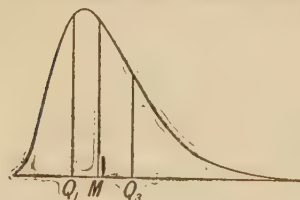
<sup>1</sup> Published by The Macmillan Company.

case the histogram will look somewhat like the surface enclosed by the upper curve at *B* in Figure 26, and the corresponding line of diagonals will look somewhat like the upper curve at *A* in Figure 26. This distribution is said to be skewed downward, or if viewed resting on its base line is said to be skewed to the left.<sup>1</sup> A skewed distribution, of course, is not a normal distribution, since a normal distribution is symmetrical.

On the other hand, if the median score of another group on the same test is close to the lower limit of scores (zero), the histogram will resemble the surface enclosed by the lower curve at *B* and the line of diagonals will resemble the lower curve at *A*. This was the case to some extent in the distribution of scores of Grade 4B, as shown in Figure 28. In this case the distribution is said to be skewed upward or to the right.

<sup>1</sup> **Measurement of skewness.** It may seem contrary to appearance to call the upper surface skewed downward and the lower surface skewed upward, and it would be more in accord with the dictionary meaning of *skewed* to say that the upper surface was skewed upward (or to the right if viewed as resting on the base line).

It happens that the various measures of skewness have been defined by Karl Pearson and others in such a way that the measure of skewness of the upper surface comes out negative, which must be interpreted as a downward skew. Similarly the measure of skewness of the lower surface comes out positive and must be interpreted as an upward skew.



Various measures of skewness have been proposed, but the one recommended is the amount of deviation of the average between the upper and the lower quartile from the median when this deviation is expressed in terms of *Q*; i.e., divided by *Q*. (See page 34 for meaning of *Q*.) That is,

$$\text{Amount of skew} = \frac{\left(\frac{Q_3 + Q_1}{2}\right) - Mdn}{Q}$$

Multiplying both numerator and denominator by 2, and remembering that  $2Q = \text{Interquartile range (see page 86)} = Q_3 - Q_1$ ,

$$\text{Amount of skew} = \frac{Q_3 + Q_1 - 2 Mdn}{Q_3 - Q_1}$$

## CHAPTER SEVEN

### PERCENTILE CURVES

**The nature of a sample.** Let us consider for a moment the scores of the 523 college freshmen represented in Figure 19. Suppose that we selected about 100 of these at random. Obviously, we should expect the median score of this sample to be about the same as the median of the whole group. If the sample is a purely *random sample*, we have no reason to expect the median to be greater or less than that of the whole group except by chance. And if it came out below in the first 100, it might be above in the second 100. The same is true of the upper and lower quartiles and of any other percentiles.

In other words, we should expect the general position and form of the distribution of a sample of a certain population to be about the same as those of the whole population.

We should expect a random sampling of ten-year children to give us a distribution of scores such that the line of diagonals would occupy the same general position and take the same general shape as the line of diagonals representing the distribution of scores of all ten-year children. We know this latter line would be a very smooth curve if no grouping were used. Therefore, if we wished to obtain the most probable position and shape of the line of diagonals of all ten-year children, we should draw a *smooth curve* in as nearly as possible the position of the line of diagonals of the distribution of scores of our sampling of ten-year children.

Now let us suppose that a certain group of pupils had taken a certain test over and over and were not able to profit by previous knowledge of it. The distribution of all the scores of these pupils would certainly represent very accurately the ability of the group in that test. We know that

the line of diagonals representing this distribution would be very nearly a smooth curve. At least the points marking the ends of the diagonals would lie in a smooth curve because of the large number of scores. Our problem, therefore, is to draw a curve in the most probable position of this smooth curve by means of the data we have, which may be only one score from each child.

**Effect of smoothing a line of diagonals.** Before discussing the method of drawing such a smooth curve, let us consider for a moment the effect that smoothing a line of diagonals has upon the corresponding histogram. In Figure 27 is shown the effect of smoothing the line of diagonals shown in Figure 18. First the points were plotted marking the ends of the diagonals, and then a smooth curve was drawn through these dots.

By *smooth curve* is meant one without kinks or humps. In that sense it is not possible, of course, to draw a smooth

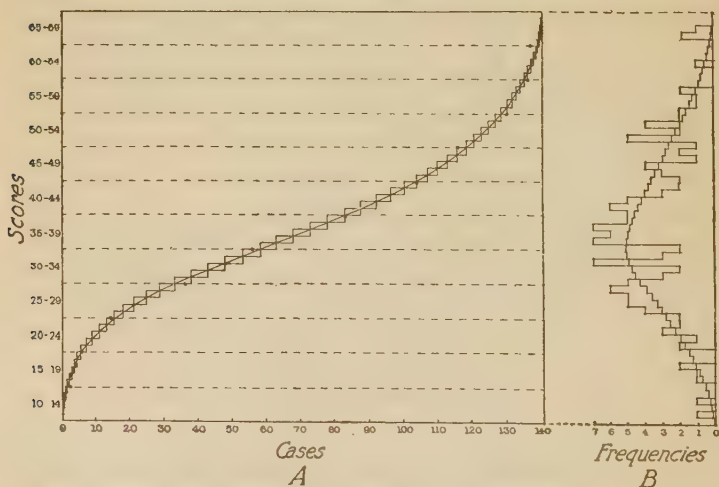


FIG. 27. Showing the effect upon a histogram of drawing a smooth curve instead of an irregular line of diagonals.

curve that will pass through all the plotted points in this case, since a curve so drawn would be somewhat wavy. The method that was followed in drawing the smooth curve through the plotted points in Figure 27 is described below.

In Figure 27 the step graph with a rectangle for each score has also been constructed from the smooth curve, and it will be seen that beginning at the bottom each little rectangle is slightly longer than the one below it up to about the middle, where the rectangles begin to shorten and do so continuously to the upper end.

These gradually increasing and decreasing rectangles are combined into the regular histogram shown at the right of the figure. They are magnified in the histogram to five times their lengths in the step graph in order better to show their relative lengths. The original histogram is also shown, and by comparing this figure with Figure 18 it may be seen how the smooth curve operates to smooth the histogram considerably more than the line of diagonals does. The height of any one of the rectangles of the smoothed histogram in Figure 27 represents theoretically the most probable frequency of that score that would occur if we tested the same group of students again under exactly the same conditions.

The rectangles in the step graph and smoothed histogram no longer represent whole numbers of scores, of course, and must be considered as representing fractional scores. Thus, the rectangles representing scores of 25, 26, 27, 28, and 29 might be approximately 2.8, 3.0, 3.3, 3.6, and 3.9 units long. The meaning of this is that if there had been ten times as many scores in the distribution (ten times as many individuals or ten scores from each), and if the distribution still retained the same general shape, we might expect the rectangles to be 28, 30, 33, 36, and 39 units long.

Whenever a given distribution of scores is considered as *representing* a larger distribution (as when a sample of un-

selected<sup>1</sup> ten-year pupils is considered as representing all ten-year pupils, or when the scores of a group of pupils in a single test represent their scores in that test repeated many times), it is proper to smooth the line of diagonals, or rather to draw a smooth curve through the plotted points in place of the line of diagonals. The median, quartiles, and other percentiles may then be found from this curve in exactly the same way as is done with the line of diagonals. Smoothing of this sort should always be done, for example, in finding norms, as will be explained.

**Percentile curves.** When a smooth curve such as that in Figure 27 is drawn through the plotted points in a percentile graph, it is called a *percentile curve*. For most practical purposes it is just as well to draw a smooth curve rather than simply to join the points by straight lines, and thus take the kinks and jogs out of the line of diagonals. The kinks and jogs in a line of diagonals are nearly always the result of pure chance; a second set of scores in the same test by the same group of pupils is quite sure to have an entirely new set of kinks and jogs. In other words, these irregularities are as a rule wholly irrelevant and accidental and may as well be smoothed out in the first place.

**The difficulty of smoothing a histogram.** You might ask whether we may not smooth the histogram in Figure 27 directly; that is, without drawing a smooth curve in place of a line of diagonals. This is impossible, except very roughly, for the reason that we cannot gauge the drawing of the smooth line through the histogram in such a way that the area included will be just equal to the area of the very irregular histogram. If it is not just equal in area to the histogram, it necessarily represents a different number of cases, which,

<sup>1</sup>The term "unselected" means a random sampling of all classes free from the effects of any selective process resulting in the choice of bright or dull pupils or advanced or retarded pupils, such as would be the case if the ten-year pupils were all taken from the fifth grade.

of course, will not do. We must keep the number of cases constant even if we do modify the frequencies of scores somewhat.

**Advantages of the percentile curve.** Now the unique feature of the smooth curve drawn in place of the line of diagonals is that it must necessarily represent exactly the same number of cases, since it is drawn within the same horizontal limits. A slight modification of the histogram is almost sure to alter the number of cases represented, whereas a variation in the position of the curve cannot in any way affect the number of cases represented.

We shall therefore discontinue the use of lines of diagonals and step graphs; these are concepts that have served a very useful purpose in enabling us to understand more fully the meaning and significance of a percentile curve. A percentile curve serves all the purposes of a step graph, line of diagonals, bar graph, histogram, or rank order, and serves most of them better.<sup>1</sup> Henceforth we shall deal almost exclusively with percentile curves. *The reader who experiences difficulty in understanding percentile curves or in drawing them, however, may continue to use lines of diagonals.* No harm can come from this.

**How to draw a percentile curve.** Let us use as an illustration the distribution of scores of Grade 5B in the mental-ability test, as shown in Table 11 and as represented in Figure 17. Figure 28 shows the same points plotted as in Figure 17, but instead of these being connected by straight lines, a smooth curve is drawn through them. By "through" of

<sup>1</sup> There is perhaps one exception to this statement. A line of diagonals is definite in the sense that any two persons using the same data should obtain exactly similar lines of diagonals, whereas two experienced persons using the same data might draw slightly differing smooth curves. It may be said, however, that the smooth curve that any person will draw according to the directions that follow will represent the distribution probably better than the line of diagonals. A slight loss in definiteness accompanies a gain in value of the graph.

UNIVERSAL PERCENTILE GRAPH

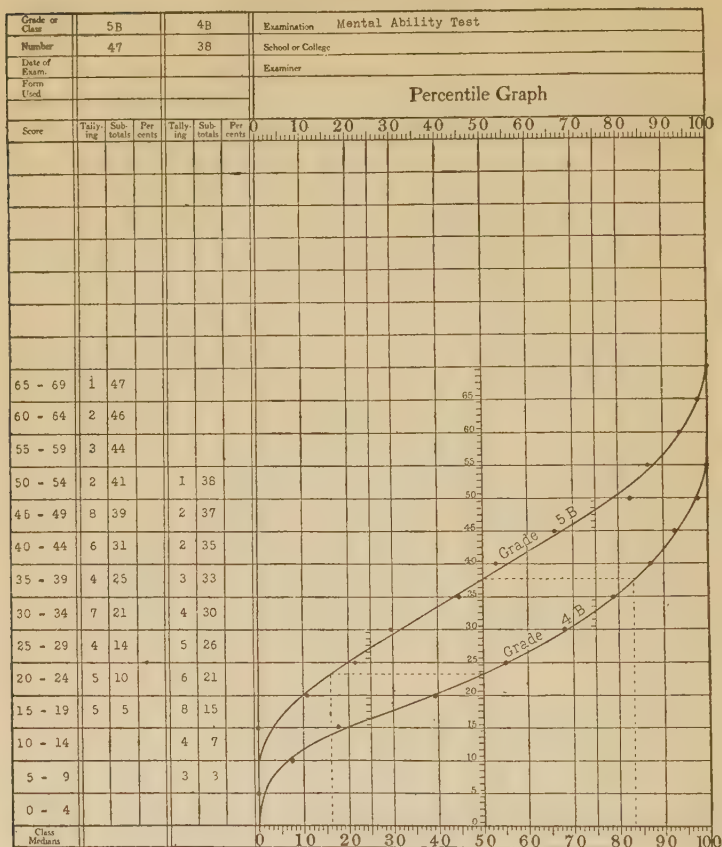


FIG. 28. Illustrating the method of drawing a percentile curve.

course we mean through or between, as the case may be. In this case our curve does not touch more than one or two of the points.

The smooth curve should be drawn so that each point above the line is balanced by one, close by, about the same

distance below the line, or in other words so that the sum of the distances of the scattered points above the line is the same as the sum of the distances of the points below the line and those distances <sup>1</sup> as small as possible consistent with the drawing of a smooth curve.

In Figure 28 the percentile curve for Grade 5B was begun below the lowest point in order that it should have the characteristic downward curve. It is probably more or less accidental that there should be five scores in the lowest frequency. We know that if the same group were tested again (neglecting practice effect, of course), it is probable that one or two of these five would drop into the interval below.

Although there is a good deal of irregularity to the points themselves, they are so located as to suggest a tendency for the curve to be approximately straight throughout its middle half; hence it was so drawn.

It is assumed that the first point is about balanced by the second, the third by the fourth, the fifth by the sixth, and the eighth by the ninth.

In the case of Grade 4B, on the other hand, the positions of the points indicate that the curve should bow downward slightly in the middle. Here the first point is balanced by the second, the third by the fourth, and the ninth by the tenth. More of the points happen to be so located that the smooth curve can be drawn directly through them.

You will see that the median score of Grade 5B as determined by the point where the curve cuts the 50-percentile line is 37, the same value that we obtained by the line of diagonals (Fig. 17). The 25- and 75-percentile lines are also cut by the curve at about the same heights that they were

<sup>1</sup> Strictly, perhaps these distances should be considered as vertical distances; but inasmuch as the deviation of points near the ends of the curve is of less consequence than the deviation of points near the middle if we think of the distances simply as perpendicular to the curve, this difference in importance is automatically taken account of.

cut by the line of diagonals. The 80-percentile line, however, is cut by the curve at a point representing a score of 50, whereas it is cut by the line of diagonals at a point representing the score of 49. Which of these values best represents the 80-percentile score of the group?

Obviously the difference is caused by the jog in the line of diagonals, and since this is presumably wholly *fortuitous* (accidental), we assume the *smoothed value* of the 80-percentile score to be the more significant one. This point merely shows the value of smoothing from one more angle. To distinguish between a median obtained by a smooth curve from the actual median, we may speak of the former as the *median by smoothing*, and the same for other percentiles. Ordinarily, however, we never obtain the actual median or the actual value of other percentiles. Usually, then, when we speak of the median of a distribution we mean the median by smoothing, and the same for the other percentiles.

EXERCISE 19. Draw on a single percentile graph the percentile curves representing scores in Grades 5A and 6A.

(For further practice see Appendix III.)

#### QUESTIONS

1. Under what circumstances would you consider a line graph as shown in Figure 17 preferable to a percentile curve as shown in Figure 28 for representing a distribution of scores, and under what circumstances would you consider the percentile curve preferable?

2. Which do you consider the more accurate as representing a distribution of scores, the actual median score of a distribution found by consulting the original papers or the median found by a carefully drawn percentile curve?

## CHAPTER EIGHT

### VARIABILITY OF SCORES

**Need for measuring of variability.** It has been found much easier to teach a class of pupils who are of approximately the same degree of mental ability than a class that contains pupils some of whom are fairly mature mentally and some immature, even though they may be of about the same age, since in the latter case some of the pupils require much more explanation than others.

Let us suppose that a principal has given a mental-ability test to the pupils of his school and has made some reclassification on the basis of the scores. (The methods employed for this purpose are explained in Chapter XXI.) The principal may wish to determine from the distributions of scores of the various classes, which teachers have the most homogeneous classes <sup>1</sup> and which the most heterogeneous.<sup>2</sup>

It would not be satisfactory merely to look at the distributions and say that this class appears to be quite heterogeneous, this one not so much so, etc. You can see the need for expressing the *variability* (relative difference between pupils in ability) of each class in terms of some single number so that we can say that the variability of this class is 20 points, of this one 18, of this one 12, etc.

There are several methods of expressing the variability of a distribution of scores in a single number. Some common measures of variability are the interquartile range, the median deviation, the average deviation, and the standard deviation. These are described in this chapter.

**The total range.** To be sure, the range of scores — that is, the interval between the lowest and highest scores — is one indication of the variability of the scores, but this measure

<sup>1</sup> Pupils alike in mental ability.

<sup>2</sup> Pupils differing in mental ability.

of variability is not a satisfactory one. In the first place we have lost track of the exact value of the scores of a group when distributing them and cannot say, for example, whether the range of scores of Grade 5B is from 15 to 69 or from 19 to 65. Moreover, if all the scores of Grade 5B had been the same except one of the three in the interval 15-19, and this one score had fallen, say, in the interval 5-9, the range would appear as 10 points greater — all the difference being caused by a change of one score.

**The interquartile range.** As explained above, the points at which a percentile curve cuts the 25- and 75-percentile lines represent what are called the *lower* and *upper quartiles* of a distribution. The interval between these points is called the *interquartile range* (I.Q.R.). For example, in Figure 28 the 5B percentile curve cuts the 25-percentile line at a height representing a score of 27 and cuts the 75-percentile line at a height representing a score of 48. The interval between these two scores, which is an interval of 21 points, is the interquartile range of this distribution.

*Problem:* Find the interquartile range of Grade 4B. *Solution:* The 25- and 75-percentile scores of Grade 4B as indicated by the heights at which the curve cuts the 25- and 75-percentile lines are, respectively, 16 and 32. The difference between these scores is 16 points. Therefore the interquartile range of the distribution of 4B scores is 16 points.

It may be seen that the interquartile range is one measure of variability. That is to say, if there is an interval of 21 points between the 25- and 75-percentile scores in one grade and an interval of only 16 points on the same test between the 25- and 75-percentile scores of another grade, the scores of the first grade are, on the whole, more variable than those of the second grade.

On the other hand, the values of the 25- and 75-percentile scores depend upon the position of the curve as a whole, and

are in no sense subject to fluctuations of a single score. For that reason the interquartile range is considered to be a far better measure of the variability of the scores of a distribution than the whole range of scores.

An equally good measure of variability is the range from the 10-percentile score to the 90-percentile score.<sup>1</sup>

**EXERCISE 20.** Find the interquartile ranges of the scores of Grades 5A and 6A from the percentile curves you drew for Exercise 19. Find also the 10-90-percentile ranges.

**Overlapping of classes.** It will be seen by a glance at the percentile curve in Figure 28 that the scores in Grade 5B are on the whole appreciably higher than those of Grade 4B, but that the distributions of scores of the two grades overlap very markedly. A convenient way to express this overlapping is to state the per cent of scores of Grade 4B that exceed the median score of Grade 5B, or to state the per cent of scores in Grade 5B that fall below the median score of Grade 4B. Thus, by finding the point on the 4B curve having a height representing a score of 37 (the median score of Grade 5B) we find that the upper 17 per cent of the scores of Grade 4B, as indicated by the curve, are above the median score of Grade 5B. The dotted lines indicate the solution.

*Problem:* Find the per cent of scores of Grade 5B that fall below the median of Grade 4B. *Solution:* The median score of Grade 4B is 23. We find the point on the 5B curve having a height of 23. This point would lie on a 16-percentile line, which indicates that 16 per cent of the pupils of Grade 5B score below the median of Grade 4B.<sup>2</sup>

<sup>1</sup> Dr. Truman L. Kelley has shown this to be more reliable than any of the more common measures of variability, which are described in this chapter.

<sup>2</sup> This does not mean necessarily that exactly 17 per cent of the actual scores of Grade 5B are below the actual median of Grade 4B, but according to the curves the score of 23 best represents the 17-percentile score of Grade 5B and the score of 23 best represents the median of Grade 4B. Therefore

**EXERCISE 21.** Find the amount of overlapping of the scores of Grades 5A and 6A in the two ways explained above. (See Exercise 19.)

**Other measures of variability ( $Q$ ).** We have already discussed the *interquartile range* as a measure of variability of a distribution. There are various other measures of variability commonly used. One of these is the *semi-interquartile range*, which is simply one half of the interquartile range. This is sometimes called simply  $Q$ . Three other measures of variability are the *median deviation*, the *average deviation*, and the *standard deviation*. Of these only the standard deviation is important, and if desired, the reader may skip the discussion of the first two. These are given, however, for reference and as a help in understanding the standard deviation.

**Median deviation.** This term refers to the median value of the deviations of the scores of a group from the median score of that group. For example, let us suppose that the scores of a small group of pupils are

4, 5, 6, 7, 8, 9, 10, 11, 12.

The median of these is, of course, 8. The score of 4 deviates 4 points from this median score of 8. The score of 5 deviates 3 points from it, etc. The deviations of the 9 scores from the median are, respectively,

4, 3, 2, 1, 0, 1, 2, 3, 4.

Now the median value of these deviations is found in exactly the same way that we should find the median of any other series of values. That is, we first put the values in order of magnitude. These then appear as follows:

0, 1, 1, 2, 2, 3, 3, 4, 4.

we are justified in expressing the overlapping of the two grades by saying that 17 per cent of the scores of Grade 5B fall below the median of Grade 4B. Strictly speaking, we should say perhaps that *if the scores in these two grades had been distributed regularly rather than irregularly*, it would then be true that 17 per cent of the scores of Grade 5B would fall below the median score of Grade 4B.

The median of these values thus arranged in order is, of course, 2. Therefore we may say that the median deviation of these scores from the median of the scores themselves is 2 points.

*Problem:* Take another series of values :

1, 3, 4, 5, 7, 8, 10, 12, 13.

Find the median deviation of these values. *Solution:* The median of these nine scores is 7. The deviations of these nine scores from the median score are, respectively,

6, 4, 3, 2, 0, 1, 3, 5, 6.

These deviations, arranged in order of magnitude, are as follows :

0, 1, 2, 3, 3, 4, 5, 6, 6.

The median of these deviations is 3.

The median deviation of the first distribution was 2 and the second 3. We may say, therefore, that the second distribution of scores is more variable than the first — it is one and a half times as variable.

**EXERCISE 22.** Find the median deviation of the distributions of scores of Grades 5A and 5B in the arithmetic reasoning test from the data given in Table 2, page 14.

**Probable error.** You will sometimes find the term *probable error* (P.E.) used to mean median deviation. There is a certain restricted sense in which this is proper, and in other cases it is improper and misleading.

If we had a hundred measures of a single magnitude, such as a hundred ratings of the handwriting of an individual, we might consider the median of these ratings as the true measure of the ability of the individual in handwriting, and in that case we should be forced to consider the other measures as in error by various amounts, plus or minus. The median value of these errors — that is, the median error (which is the same as the median deviation of the distribution of measures) —

may be properly called the probable error in the sense that the chances are even (one to one) that the error of any given measure will not exceed that amount.

The median deviation of the distribution of a series of measures of a single magnitude (when there are enough measures so that the median measure may be considered the true measure) is, therefore, properly called the probable error (P.E.) of the measures. In all other cases it is improper to call the median deviation the probable error.

**Average deviation** (or **Mean deviation**).<sup>1</sup> This term refers to the average of the deviations of scores of a group from the average of those scores. (Note that the median deviation is the median of deviations measured from the *median* score, whereas the average deviation is the average of deviations measured from the *average* score.)

For example, let us take the same distribution :

4, 5, 6, 7, 8, 9, 10, 11, 12.

The sum of these scores is 72 and the average is 8, the same as the median. The deviations of these values from the average, 8, are as before :

4, 3, 2, 1, 0, 1, 2, 3, 4.

We have now to find the average of these deviations. Their sum is 20. Dividing 20 by 9, the number of cases, our average deviation comes out  $2\frac{2}{9}$ , or approximately 2.22. (Note that the deviation of 0 counts the same as any other deviation.) This, you will note, is not exactly the same as the value we got for the median deviation, but the median deviation was the whole number nearest to this value.

<sup>1</sup> The term *average* is considered by certain authorities as referring to any measure of central tendency, while the term *mean* refers to what is commonly known as the average. In this case, however, it is convenient to use the term *average deviation*, since the abbreviation Avg. Dev. (sometimes written A D.) is not likely to be confused with Med. Dev. (sometimes written M.D.). Since M.D. might be understood as standing for mean deviation, it is safer to write Med. Dev.

*Problem:* Find similarly the average deviation of this series of values:

1, 3, 4, 5, 7, 8, 10, 12, 13.

*Solution:* The sum of these values is 63. Dividing this by 9, the number of cases, we get 7 as the average score (this being again the same as the median). The deviations of the 9 scores from this average are, respectively,

6, 4, 3, 2, 0, 1, 3, 5, 6.

The sum of these 9 deviations is 30. Dividing 30 by 9, we get  $3\frac{2}{3}$ , or approximately 3.33, as the average deviation. Here again the average is a little higher than the median. There is a tendency for the average deviation of a distribution of scores to come out a little higher than the median deviation. However, the ratio of the average deviations of two distributions tends to be the same as the ratio of the median deviations. For example, the ratio of the median deviations of these two hypothetical distributions we find to be 2:3, and you will note that the ratio of the deviations is the same.

**EXERCISE 23.** Find the average deviations of the distributions of scores of Grades 5A and 5B in the arithmetic reasoning test. (See Table 2.) Compare these with the median deviations found in Exercise 22.

**Standard deviation.** The standard deviation is a little more complicated than any of the preceding measures of variability, but it is considered by statisticians to be somewhat more valid. It is used in those cases where very great accuracy is necessary or in which, for one reason or another, it is a little more convenient to determine than any of the other measures of variability. For all ordinary purposes of the teacher or principal, however, the interquartile range found by means of a percentile curve is accurate enough. The method of finding a standard deviation is given here, however, for reference.

The standard deviation of a distribution is the square root of the mean of the squares of the deviations of scores from the mean of the scores. Although this may sound somewhat confusing, an example illustrating the calculation of a standard deviation will make the method perfectly clear.

The calculation of a standard deviation of scores in the two hypothetical distributions which we have been considering is given in Table 12.

TABLE 12

FIRST HYPOTHETICAL DISTRIBUTION			SECOND HYPOTHETICAL DISTRIBUTION		
Scores	Deviations from Mean	Deviations Squared	Scores	Deviations from Mean	Deviations Squared
4	4	16	1	6	36
5	3	9	3	4	16
6	2	4	4	3	9
7	1	1	5	2	4
8	0	0	7	0	0
9	1	1	8	1	1
10	2	4	10	3	9
11	3	9	12	5	25
12	4	16	13	6	36
		60			136
$60 \div 9 = 6.67$ $\sqrt{6.67} = 2.58$ 2.58 = standard deviation of this distribution			$136 \div 9 = 15.11$ $\sqrt{15.11} = 3.89$ 3.89 = standard deviation of this distribution		

Under "Scores" in the left-hand column are the scores in the first hypothetical distribution. In the next column appear the deviations of these scores from the average (8) of the scores. In the next column appear the squares of these deviations. Thus, beginning at the top of the columns,  $4^2$  equals 16,  $3^2$  equals 9, etc. According to the definition of a standard deviation given above, we must find the average of these

squares of the deviations. The sum of the squares of the deviations is shown to be 60. The symbol for the sum of the squares of a series of deviations is  $\Sigma d^2$  ("sum of  $d$  square") ( $\Sigma$  = large Greek letter  $S$  and is read "sum of"). Dividing this sum by 9, we find the average of these squares of the deviations to be 6.67. The square root of this average is 2.58. The value 2.58 is the standard deviation of the distribution of scores. The symbol for the standard deviation of a distribution is  $\sigma$  (small Greek letter  $s$ , called sigma).

**Variability of normal distributions.** In a theoretically normal distribution there is a certain fixed relation between the interquartile range, the median deviation, the average deviation, and the standard deviation. These bear approximately the following relations:

$$\begin{aligned} \text{I.Q.R.} &= 2 \text{ Med. Dev.} \\ \text{Med. Dev.} &= .67 \sigma \\ \text{Med. Dev.} &= .85 \text{ Avg. Dev.} \\ \text{Avg. Dev.} &= .80 \sigma \\ \text{Avg. Dev.} &= 1.18 \text{ Med. Dev.} \\ \sigma &= 1.25 \text{ Avg. Dev.}^1 \\ \sigma &= 1.48 \text{ Med. Dev.} \end{aligned}$$

Figure 29 shows these relations graphically.

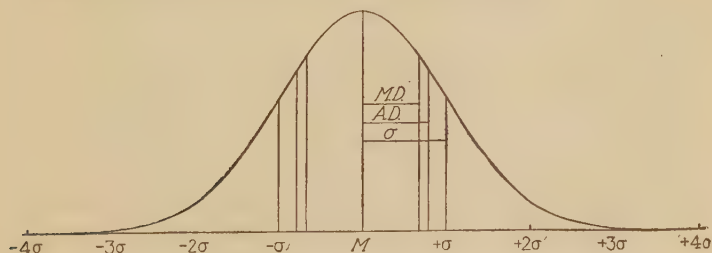


FIG. 29. Showing the relative values of the median deviation (M.D.) (sometimes called the probable error, P.E.), the average deviation (A.D.), and the standard deviation ( $\sigma$ ) of a normal distribution.

<sup>1</sup> More precisely,  $\sigma = 1.25331$  Avg. Dev. or  $1.48256$  Med. Dev.

These equations are often used to convert one measure of variability into terms of another. For example, let us suppose that we have had occasion to find the standard deviation of a distribution of scores and wish to know what the median deviation of the distribution would probably be. If the distribution is approximately normal, it is proper to assume that the median deviation would most probably be about .67 times the standard deviation. Suppose that in our distribution  $\sigma = 25$ . Then the most probable value of Med. Dev. =  $.67 \times 25 = 17$ .

In the same way, of course, one may find the most probable value of the standard deviation of a distribution from the median deviation or average deviation, etc.

Statisticians who feel obliged to secure very great accuracy use the formula Med. Dev. =  $.6745 \sigma$ , and indeed it is sometimes felt that a value of Med. Dev. so obtained is a better measure of median deviation than the actual median deviation, just as a median by smoothing is a better median than the actual median.

**EXERCISE 24.** The finding of the standard deviation ( $\sigma$ ) of the scores in the second hypothetical distribution is also given in Table 12. Cover up all figures except the scores in the column under "Scores" and find the standard deviation, comparing this with the value shown. Find the ratio of the two standard deviations and compare it with the ratio of the median and average deviations.

**EXERCISE 25.** If the standard deviation of a distribution is 43, what is the most probable value of the average deviation, median deviation, and interquartile range? Use the approximate values in the table.

**EXERCISE 26.** If the I.Q.R. of a distribution is 24, what is the most probable value of the Med. Dev., Avg. Dev., and  $\sigma$ ?

### QUESTION

Do you think the value of  $Q$  in a markedly skewed distribution is greater or less than the value of the median deviation? (It is the same, you know, in a normal distribution.)

## CHAPTER NINE

### PERCENTILE RANK IN A NORMAL DISTRIBUTION

It has been more convenient to reserve until after the chapter on variability of scores the discussion of the very important relation between scores and percentile ranks which we shall now consider.

In Figure 30, on the next page, is shown a typical percentile curve. You have learned how to find the percentile rank of any score in a distribution of scores by drawing a percentile curve to represent the distribution.

**Units of percentile rank and units of score compared.** Are the units of percentile rank in the distribution represented in Figure 30 proportional to the corresponding units of score? That is, does each interval of percentile rank of, say, 10 units correspond to the same number of units of score?

You will see by examining the percentile curve that the interval in percentile rank from 50 to 60 embraces the interval of score from 62 to about 66 (4 units). The interval in percentile rank from 60 to 70 embraces the scores from about 66 to 70 (4 units). The next 10 units of percentile rank embrace the scores from about 70 to  $74\frac{1}{2}$  ( $4\frac{1}{2}$  units); the next 10 embrace scores from  $74\frac{1}{2}$  to 81 ( $6\frac{1}{2}$  units); and the last 10 units of percentile rank embrace all scores from 81 to as far up as the scores extend.

It is obvious from the above that units of percentile rank and units of score are not comparable in the sense of being proportional. The same number of units of percentile rank correspond to different amounts of score in different portions of the distribution.

**The correspondence between scores and percentile ranks.** Let us test the relation between scores and percentile ranks in the opposite manner — that is, by taking equal increments

UNIVERSAL PERCENTILE GRAPH

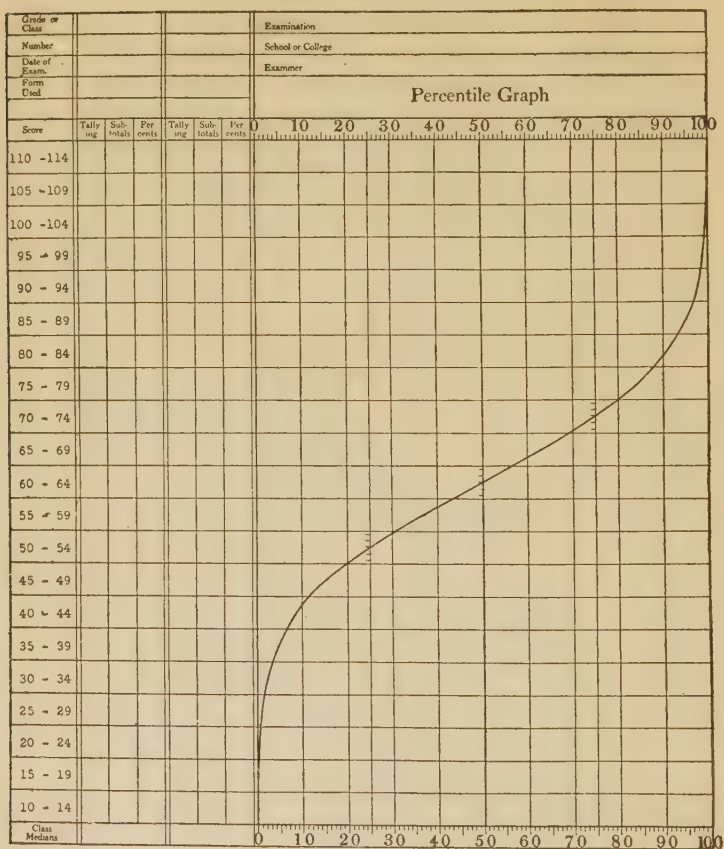


FIG. 30. A normal percentile curve.

of score. The interval of 5 units of score from the median, 62, to 67 embraces the interval of percentile rank from 50 to 63 (13 units). The next 5 units of score (67 to 72) embrace the interval of percentile rank from 63 to 75 (12 units). The next 5 units of score embrace percentile ranks from 75 to 84½

( $9\frac{1}{2}$  units); the next from  $84\frac{1}{2}$  to 91 ( $6\frac{1}{2}$  units); next from 91 to  $95\frac{1}{2}$  ( $4\frac{1}{2}$  units); etc., with ever-decreasing intervals of percentile rank.

The correspondence between scores and percentile ranks at intervals of 5 points in score for the distribution represented in Figure 30 is shown in Table 13 between the scores of 97 and 27. It should be understood that these are scores of an altogether hypothetical test, that the median was set arbitrarily at 62 for convenience, and the value of  $Q$  was arbitrarily made 10 points.

TABLE 13

SHOWING THE CORRESPONDENCE BETWEEN SCORES AND PERCENTILE RANKS IN THE HYPOTHETICAL TYPICAL DISTRIBUTION OF SCORES REPRESENTED BY THE PERCENTILE CURVE IN FIGURE 30. A SCORE OF 97 CORRESPONDS TO A P.R. OF 99, ETC.

SCORE	P.R.	SCORE	P.R.	SCORE	P.R.
97	99	72	75	47	$15\frac{1}{2}$
92	98	67	63	42	9
87	$95\frac{1}{2}$	62	50	37	$4\frac{1}{2}$
82	91	57	37	32	2
77	$84\frac{1}{2}$	52	25	27	1

Comparing scores and percentile ranks in this way, we see also that units of percentile rank in the extremes embrace much larger amounts of score than in the middle portion of the distribution. The farther one goes toward the extremes, the greater the amounts of score corresponding to each unit of percentile rank.

This same tendency is found in the case of all distributions, no matter in what portion of the scale of scores the distribution is situated. For that reason it would appear that the variation in unit value is in the percentile ranks rather than in the scores. There may be some variation in value of units of score from one part of a scale to another, of course, but this

in no way accounts for the increasing value of units of percentile rank as we approach the extremes of a distribution.

**Percentile rank in terms of variability.** You will see by the percentile curve in Figure 30 and by Table 13 that the interquartile range of our hypothetical distribution is 20 points (52 to 72). And since the distribution is symmetrical, with the median, 62, just halfway between the upper- and lower-quartile scores, the median deviation of the distribution is 10 points. If we substitute Median for the score of 62 (Median + Med. Dev.), for the score of 72, etc., in Table 13, we have Table 14.

TABLE 14

SCORE	P.R.	SCORE	P.R.	SCORE	P.R.
(M. + $3\frac{1}{2}$ M.D.)	99	(M. + M.D.)	75	(M. - $1\frac{1}{2}$ M.D.)	$15\frac{1}{2}$
(M. + 3 M.D.)	98	(M. + $\frac{1}{2}$ M.D.)	63	(M. - 2 M.D.)	9
(M. + $2\frac{1}{2}$ M.D.)	$95\frac{1}{2}$	Median	50	(M. - $2\frac{1}{2}$ M.D.)	$4\frac{1}{2}$
(M. + 2 M.D.)	91	(M. - $\frac{1}{2}$ M.D.)	37	(M. - 3 M.D.)	2
(M. + $1\frac{1}{2}$ M.D.)	$84\frac{1}{2}$	(M. - M.D.)	25	(M. - $3\frac{1}{2}$ M.D.)	1

Table 14 shows the approximate percentile ranks of scores of a normal distribution in terms of the median deviation of the distribution. The table is interpreted to mean that in any exactly normal distribution, no matter what the median score may be or the median deviation of the distribution, an individual making a score just one half of the median deviation above the median will have a percentile rank of 63; an individual making a score above the median by twice the amount of the median deviation will have a percentile rank of 91; a score three times the median deviation above the median corresponds to a percentile rank of 98; etc.

**Testing the normality of a distribution.** We may use Table 14 to test a distribution to determine whether it is approximately a normal distribution. Thus Dr. Terman

gives the percentile ranks of individuals who attain each of various intelligence quotients <sup>1</sup> shown in Table 15.<sup>2</sup>

TABLE 15

IQ	P.R.	IQ	P.R.
130	99	95	33 $\frac{1}{3}$
128	98	92	25
125	97	90	20
122	95	88	15
116	90	85	10
113	85	78	5
110	80	76	3
108	75	73	2
106	66 $\frac{2}{3}$	70	1
100	50		

*Problem:* Is the distribution of intelligence quotients a normal distribution? *Solution:* The median positive deviation (interval from the 50th to the 75th percentile rank) is 8 points (100 to 108), and the median negative deviation (interval from the 50th to the 25th percentile rank) is also 8 points (100 to 92). Therefore we may consider 8 points in intelligence quotient as the median deviation of the distribution of intelligence quotients. According to Table 14, therefore, we should expect the percentile rank of an individual having an IQ of (Median - 2 Med. Dev.), which would be an IQ of 84, to be 9. Since the percentile rank corresponding to an IQ of 85 is 10, it is quite reasonable to suppose that the percentile rank corresponding to an IQ of 84 would be 9. Similarly we should expect the percentile rank of an individual having an intelligence quotient of (Median - 3 Med. Dev.), which would be an IQ of 76, to be 2. Dr. Terman's data

<sup>1</sup> For the meaning of intelligence quotient, see Chapter XIII, "The Measurement of Brightness."

<sup>2</sup> L. M. Terman, *The Measurement of Intelligence*, page 78; Houghton Mifflin Company.

show the percentile rank of an individual having an IQ of 76 to be 3 instead of 2. This slight deviation may be caused by the true median deviation being slightly more than 8 points, although given as 8 as the nearest whole number. A similar comparison of IQ's and percentile ranks above the median shows a fairly close agreement with the correspondence that would be expected in the case of a normal distribution.

On the other hand, if we tested the distribution of scores of Grade 4B in the mental-ability test shown in Figure 28 for normality, we should find quite a divergence. First of all, the positive median deviation is 9 points ( $32 - 23$ ), while the negative median deviation is only 7 points ( $23 - 16$ ). If we took 8 points as the median deviation of the distribution, we should expect a pupil making a score of (Median + 2 Med. Dev.) ( $23 + 16 = 39$ ) to have a percentile rank of 91. According to the percentile curve, however, a score of 39 corresponds to a percentile rank of only 86. These discrepancies and others that would be found show that the distribution of scores in this case was not a normal distribution.

**An accurate table of percentile ranks.** In Appendix II (page 294) is given a more complete table similar to Table 14 showing the exact percentile ranks of individuals making scores of (Median + .1 Med. Dev.), (Median + .2 Med. Dev.), etc., in the case of a normal distribution.

## CHAPTER TEN

### THE CORRESPONDENCE BETWEEN TESTS

**The problem.** Let us suppose that we have given the National Intelligence Test and the Otis Self-Administering Tests, Intermediate Examination, to the same pupils and wish to compare the scores of various pupils in the two tests to determine whether they did better in the National Intelligence Test or in the Intermediate Examination.

It is obvious, of course, that a score of 50 in the National Intelligence Test does not necessarily represent the same ability that a score of 50 in the Intermediate Examination represents. First of all, there are more items in the National Intelligence Test and this might enable the pupil to make a larger score. Moreover, the items in the Intermediate Examination may be on the whole somewhat more difficult, so that the pupils could not cover as many in the same time. Also, the time limits are not the same. Still another factor is the effect of practice. If the National Intelligence Test was given first, the pupil will do better in the Intermediate Examination than if the Intermediate Examination was given first. And if there was an appreciable lapse of time between the tests, the element of growth of the pupil in mental ability enters.

Let us say that a certain pupil made a score of 95 in the National Intelligence Test and a score of 30 in the Intermediate Examination. In order to determine, therefore, whether the pupil did better (relative to the class) in the National Intelligence Test or in the Intermediate Examination, we must know just what score in the National Intelligence Test corresponds to a score of 30 in the Intermediate Examination or what score in the Intermediate Examination corresponds to a score of 95 in the National Intelligence Test.

And the same for all the other scores in the tests. In other words, we must know either the values of the various National Intelligence Test scores in terms of scores in the Intermediate Examination, or the values of the various Intermediate Examination scores in terms of scores in the National Intelligence Test. Or, as we sometimes say, we must transmute the scores of one test into terms of scores of the other.

It should be remembered that since the correspondence between scores is determined to some extent by the conditions under which the tests were given, — which was given first, what time elapsed between them, and even perhaps the degree of familiarity of the pupils with standard tests, — the correspondence must be found in each particular case in order to express scores of one test in terms of another with reasonable assurance of accuracy.

It is a common practice to compare scores of two tests by expressing the scores of each in terms of mental age. This would be an excellent way except for the influence of practice effect, and of time elapsed between tests, and the fact that no two tests are standardized on the basis of the same set of scores and the methods of standardization are not yet themselves sufficiently standardized so that we can rely on the results with certainty.

**A method of finding the correspondence between scores in two tests.** If the number of cases on which the comparison of two tests is to be based is small, say less than 50, it is not possible to find the correspondence very accurately, and any convenient method that will yield a first approximation, so to speak, will do. One such method is that of comparing scores of the same rank in magnitude. Thus, let us take the case of our 8A class whose scores in the National Intelligence Test and Intermediate Examination are shown in Table 16.

TABLE 16

PUPIL	N.I.T. SCORE	I.E. SCORE	PUPIL	N.I.T. SCORE	I.E. SCORE	PUPIL	N.I.T. SCORE	I.E. SCORE
1	119	52	12	142	53	23	111	55
2	128	59	13	135	36	24	125	58
3	133	43	14	144	54	25	153	65
4	158	59	15	136	48	26	124	54
5	155	64	16	142	42	27	127	52
6	139	61	17	103	50	28	136	44
7	145	55	18	156	58	29	150	55
8	120	60	19	143	64	30	148	71
9	142	63	20	135	66	31	117	52
10	136	56	21	144	64	32	142	59
11	130	45	22	131	63	33	123	59

If we arrange the scores of these 33 pupils in each test in the order of magnitude, they will appear as shown in Table 17.

TABLE 17

N.I.T. SCORE	I.E. SCORE	N.I.T. SCORE	I.E. SCORE	N.I.T. SCORE	I.E. SCORE
158	71	142	59	130	53
156	66	142	59	128	52
155	65	142	59	127	52
153	64	139	58	125	50
150	64	136	58	124	48
148	64	136	56	123	45
145	63	136	55	120	44
144	63	135	55	119	43
144	61	135	55	117	42
143	60	133	54	111	42
142	59	131	54	103	36

Now if we assume that the highest score made in the National Intelligence Test is the equivalent of the highest score made in the Intermediate Examination (although by a different pupil, of course), then an N.I.T. score of 158 equals an

I.E. score of 71. If we assume that the next highest score made in the N.I.T. is the equivalent of the next highest score made in the I.E., then an N.I.T. score of 156 equals an I.E. score of 66. In the same way we may assume the equivalence of scores of the same rank throughout the ranges and say that an N.I.T. score of 155 equals an I.E. score of 65, etc.

Of course, this rough method leads one into difficulty, for the N.I.T. score of 144 is found opposite I.E. scores of 63 and 61, but in such a case we might "split the difference" and say that 144 corresponds to 62.

**A refinement of the method of equating scores.** The inconsistency just cited suggests that this method of equating scores is very rough. The inaccuracy of the method is brought out by Figure 31. Thus, in Figure 31, the vertical scale represents scores in the National Intelligence Test and

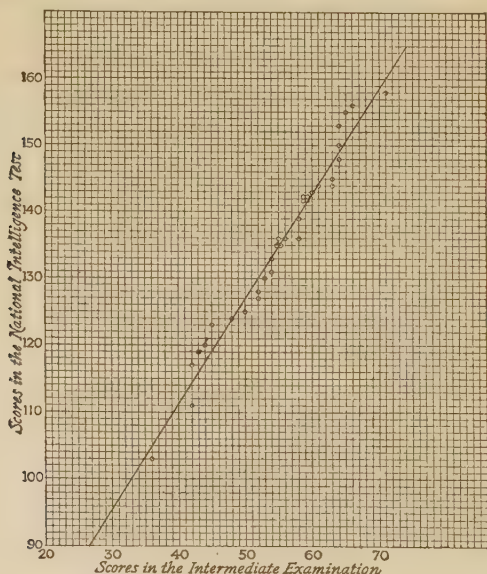


FIG. 31. Showing the method of drawing a line of relation.

the horizontal scale represents scores in the Intermediate Examination. The small circle in the upper right-hand corner represents the pair of highest scores (158 and 71), as shown in Table 17. It has a height of 158 by the vertical scale and is directly above 71 on the horizontal scale. The next circle represents the next pair of scores (156 and 66), etc.

It is plain from this figure that the correspondence between the tests as found by the method of equating ranks outlined above is indeed very rough. But having thus plotted the pairs of scores that have the same rank order, it is possible to see by inspection what the general trend of the correspondence is and to draw a straight line, as shown in the figure, in such a position as best to represent this general trend. This line is called the *line of relation* between scores in the two tests. The position of this line was determined in this case simply by the eye, in the same way that the position of a percentile curve is determined.

**Reading the line of relation.** Having drawn a line of relation between the scores in the two tests, it is possible to determine at once what score in the Intermediate Examination corresponds to any score in the National Intelligence Test, according to the line, and what score in the National Intelligence Test corresponds to any score in the Intermediate Examination.

Thus, since the line in Figure 31 cuts the horizontal line representing an N.I.T. score of 100 at a point above 33 on the scale of I.E. scores, we should say that a score of 100 in the N.I.T. corresponds in this case to a score of 33 in the I.E., and vice versa. The line cuts the 130 line at a point above 52; so we should say that the I.E. score corresponding to an N.I.T. score of 130 is 52.

To find the I.E. score corresponding to the N.I.T. score of 119 (the score of pupil No. 1 in Table 16), note the point at which the line of relation cuts the horizontal line opposite the

point 119 on the N.I.T. scale and note the point on the I.E. scale just below this point of intersection. This will be found to be 45, showing that 45 is the I.E. score that corresponds to 119 in the N.I.T.

This pupil made 52 in the I.E. ; so he did better in the I.E. than in the N.I.T.

Similarly, to find the N.I.T. score corresponding to an I.E. score, say, of 59 (the score of the second pupil in Table 16), reverse the process. Thus the line of relation cuts the vertical line 59 at a point opposite 141 on the vertical scale. Therefore, 141 is the N.I.T. score corresponding to an I.E. score of 59. This second pupil made only 128 in the N.I.T., and therefore he also did better in the I.E.

If, as in the case of the I.E. score of 43 (third pupil in Table 8), we find the line of relation cutting the vertical line (43) at a point between two horizontal lines (in this case 115 and 116), we take the nearest line, or if the intersection seems to be exactly halfway between two lines, take the one representing the even number.

**EXERCISE 27.** Find which pupils in the first column of Table 16 did better in the National Intelligence Test than in the Intermediate Examination. Check your results by transmuting first N.I.T. scores to I.E. scores, then I.E. scores to N.I.T. scores.

**Making a table of correspondence.** It is not so easy, of course, to determine correspondence by means of the line of relation as by means of a table ; and if it is desired for any reason to provide for easy finding of correspondence, a table of correspondence may be made from the graph. A table is generally made for transmuting only one way ; that is, consecutive scores in one test are given with the corresponding scores in the other, which may involve skips or duplication.

Let us suppose that it is desired to make a table for transmuting I.E. scores into terms of N.I.T. scores. We should

then begin by writing in the table the consecutive scores in the Intermediate Examination, as shown in Table 18, and opposite each one the N.I.T. score corresponding to it, as determined by the line of relation.

TABLE 18

I.E. SCORES	N.I.T. SCORES
33	100
34	102
35	103
36	105
37	106
38	108
39	110
40	111
41	113
42	114
43	116
44	118
45	119

EXERCISE 28. Complete Table 18. Make also a table for converting N.I.T. scores between 100 and 120 into I.E. scores. The table will begin thus :

N.I.T. SCORES	I.E. SCORES
100	33
101	34
102	34
103	35
104	35
105	36
106	37

USE OF PERCENTILE GRAPH IN FINDING  
CORRESPONDENCE

In the preceding pages we have considered a method for finding the correspondence between scores in two tests when only a few pupils had been tested by both tests. The results in such a case, however, are not reliable and, if possible, a much larger number of cases should be used. Several hundred cases should be used to obtain a fair degree of accuracy even for the conditions of a given investigation. Even these cannot be considered as sufficient to establish the correspondence in general between two tests. For that purpose several thousand cases are needed — a number approximating the number of cases needed to establish age and grade norms.

Let us suppose that we wish to find the correspondence between scores in the National Intelligence Test and scores in the Intermediate Examination, making use of the 284 scores in these tests made by pupils in our grades 4B to 8A. It would be very tedious to arrange the scores of the 284 pupils in both tests in exact rank orders. For that reason the method described above is not suitable in this case. A better method is described in the following pages.

If the scores of the pupils have not been distributed already for some purpose, the distributions might be found, as shown in Table 19.

A very rough idea of the correspondence may be had from inspection of this table. We can see the range of scores in each test and form a general idea of the location of the central tendencies. We can see at a glance, of course, that the scores in the National Intelligence Test tend to run much higher than those in the Intermediate Examination. If the extreme scores were significant, we might say that a score of 70 in the Intermediate Examination seems to correspond to a score of about 180 in the National Intelligence Test and

TABLE 19

NATIONAL INTELLIGENCE TEST		INTERMEDIATE EXAMINATION	
Score Interval	Frequencies	Score Interval	Frequencies
0 to 9		0 to 4	
10 to 19		5 to 9	2
20 to 29	1	10 to 14	7
30 to 39	1	15 to 19	13
40 to 49	1	20 to 24	15
50 to 59	0	25 to 29	25
60 to 69	8	30 to 34	21
70 to 79	14	35 to 39	26
80 to 89	23	40 to 44	29
90 to 99	25	45 to 49	34
100 to 109	37	50 to 54	37
110 to 119	37	55 to 59	40
120 to 129	51	60 to 64	21
130 to 139	32	65 to 69	13
140 to 149	29	70 to 74	1
150 to 159	18		
160 to 169	5		
170 to 179	1		
180 to 189	1		

a score of 5 in the Intermediate Examination seems to correspond to a score of about 20 in the National Intelligence Test. But these extreme scores are very unreliable in the determination of correspondence and at any rate would not give us the correspondence between intervening scores.

We could say with assurance that the median N.I.T. score corresponds to the median I.E. score, and similarly we may say that the upper-quartile (75-percentile) score of the one distribution corresponds to the upper-quartile score of the other, and the same for the lower-quartile scores.

The best way to find the median and upper- and lower-quartile scores of a distribution is, of course, by drawing a percentile curve. In Figures 32 and 33 are shown the percentile curves representing the two distributions of score,

UNIVERSAL PERCENTILE GRAPH

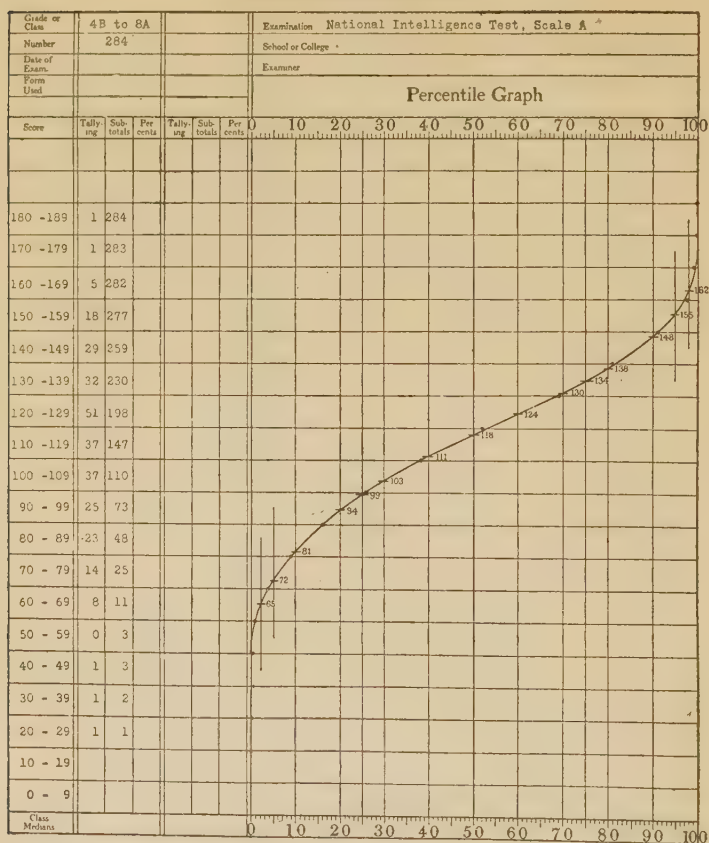


FIG. 32. A percentile graph representing the distribution of scores of the 284 pupils in Grades 4B to 8A in the National Intelligence Test, Scale A.

We are now in a position to find very easily the median scores and the upper- and lower-quartile scores of each distribution. These are indicated by the points where the 50-percentile lines and the 25- and 75-percentile lines cut the two curves. The median N.I.T. score is 118, and the median

UNIVERSAL PERCENTILE GRAPH

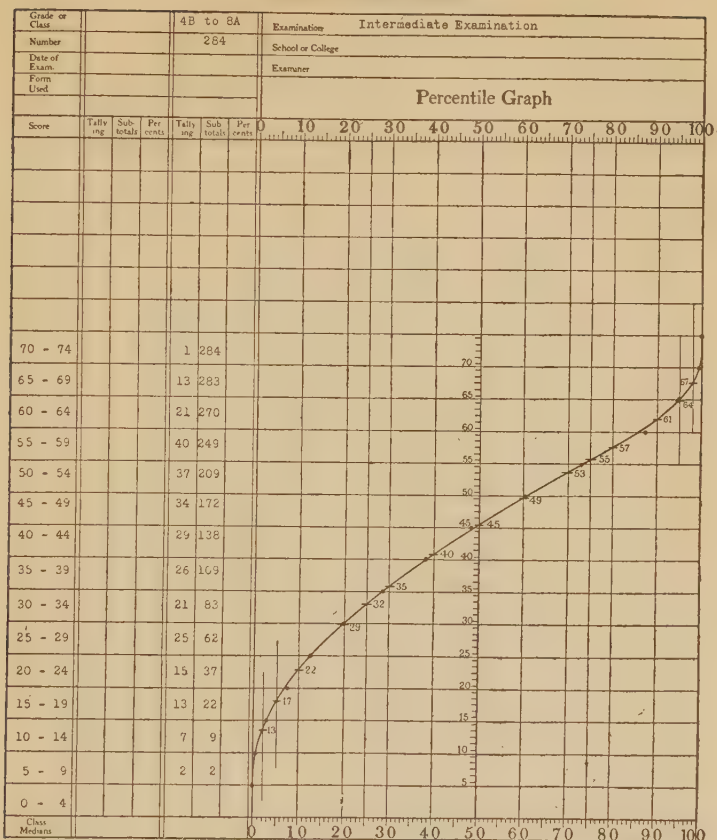


FIG. 33. A percentile curve representing the distribution of scores of the 284 pupils in Grades 4B to 8A in the Intermediate Examination.

I.E. score is 45. Therefore we may assume that an N.I.T. score of 118 corresponds to an I.E. score of 45.

The upper-quartile scores in the two tests are, respectively, 134 and 55; so we may say that an N.I.T. score of 134 corresponds to an I.E. score of 55. The lower-quartile scores in

the two tests are, respectively, 99 and 32; so we may say that an N.I.T. score of 99 corresponds to an I.E. score of 32.

In this same way we could go on and find other pairs of corresponding scores, such as the two 40-percentile scores and the two 60-percentile scores. Indeed, we could take each percentile — the 1st, 2d, 3d, etc. — and make up a table showing the correspondence between scores in the two tests by giving each pair of scores thus found.

**The line of relation.** We have already learned a better way than this to make up such a table, and that is by means of a line of relation. In this case, of course, we shall not plot actual scores having the same rank, but we shall do what amounts to exactly the same thing, we shall plot pairs of corresponding percentile scores.

It is not necessary, of course, to plot pairs of scores representing every percentile. It is usually sufficient to take only every tenth percentile between the 10th and 90th. The 0- and 100-percentiles are of practically no value, since they are so unstable. There should be added, however, the 5th and 95th and 2d and 98th percentiles, since otherwise the correspondence is not carried far enough up and down. Even the 2d and 98th percentiles, however, must be regarded as of doubtful value, since even these are very unstable.

Our next step toward drawing a line of relation, then, is to make a table containing the pairs of corresponding percentile scores in the two distributions. A table drawn up as suggested above appears as shown in Table 20.

We are now ready to plot points representing these pairs of corresponding percentile scores. Our first step, of course, is to lay out scales on cross-section paper to represent scores in the two tests as shown at the foot and left side in Figure 34. The next step is to plot a point to represent each pair

TABLE 20

SHOWING CORRESPONDING PERCENTILE SCORES IN THE INTERMEDIATE EXAMINATION AND NATIONAL INTELLIGENCE TEST SELECTED FOR PLOTTING TO DRAW A LINE OF RELATION BETWEEN SCORES IN THE TWO TESTS

PERCENTILE	I.E.	N.I.T.	PERCENTILE	I.E.	N.I.T.
98	67	162	40	40	111
95	64	155	30	35	103
90	61	148	20	29	94
80	57	138	10	22	81
70	53	130	5	17	72
60	49	124	2	13	65
50	45	118			

of scores in our table. Thus in Figure 34 the uppermost point represents the pair of 98-percentile scores, 67 and 162; the next point represents the pair of 95-percentile scores, 64 and 155, etc., the lowest point representing the pair of 2-percentile scores.

You will note that the points lie in such a way that a smooth line may be drawn passing through practically every point. The line is somewhat bent, but its position seems quite definite, at least within the range of points plotted.

You will note a general tendency for lines of relation to be straight, although in various instances there will be deviations from this rule.

The line of relation in Figure 34 indicates that toward the lower ends of the distributions scores in the two tests are probably so related that an interval of 10 points in the Intermediate Examination corresponds to an interval of about 18 points in the National Intelligence Test. Near the middle of the distribution the correspondence is such that an interval of 10 points in the Intermediate Examination is the equivalent of an interval of 14 or 15 points, and in the upper portions 10 points in the Intermediate Examina-

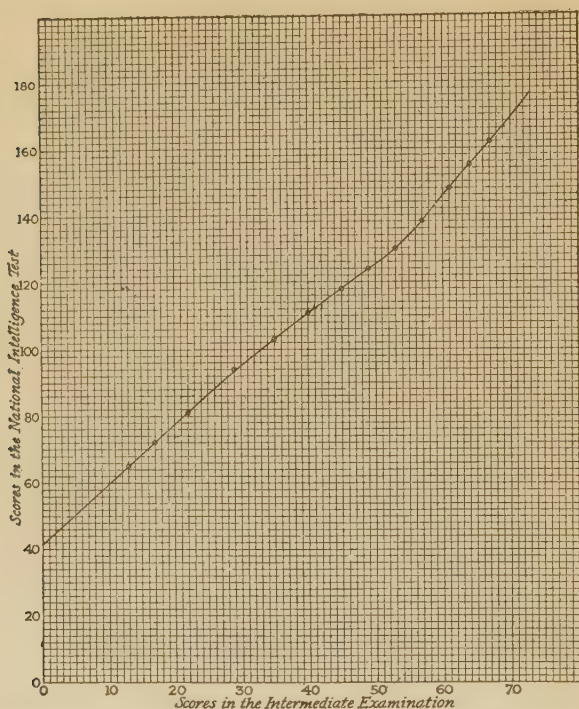


FIG. 34. Showing the method of finding a line of relation between scores in two tests by plotting pairs of percentile scores obtained by percentile graphs. (See Figures 32 and 33.)

tion correspond to 25 or 26 points in the National Intelligence Test.

As has been explained, the extreme scores are very unreliable as indications of correspondence. Indeed, it is hardly worth while to consider any percentile scores above the 98-percentile score or below the 2-percentile score. We may, if we wish, extend our straight line outside the limits of our 2- and 98-percentile points; but the correspondence between scores outside these limits must be considered as

very indefinite, and scores in one test converted into terms of the other outside these limits must be considered as of very doubtful value.

Thus it is seen that according to the line of relation a score of 0 in the Intermediate Examination corresponds to a score of about 42 in the National Intelligence Test. Now it is hardly conceivable that a pupil who would miss such questions as "What is the opposite of west?" in the Intermediate Examination could answer 42 questions in the National Intelligence Test. It is probable, therefore, that the true line of relation would bend downward somewhat in the lower extreme. A downward bend of the line of relation in the lower portion would indicate that there were so many more easy questions in the National Intelligence Test than in the Intermediate Examination that a difference of 10 points in the Intermediate Examination in the lower range corresponds to a difference of somewhat *more* than 18 points in the National Intelligence Test.

As has been said, the fewer the scores used in finding the line of relation, the less accurate the position of the line will be. In this case, we had 284 scores, which are probably enough to establish the position of the line with a fair degree of accuracy for the particular survey from which these scores were obtained.

**Drawing up the table of correspondence.** After having drawn a line of relation between the National Intelligence Test and the Intermediate Examination, we are in a position to draw up a table of correspondence between the two tests based on the 284 cases. Let us suppose that it is desired to transmute Intermediate Examination scores into terms of the National Intelligence Test. We should find from the line of relation in Figure 34 the N.I.T. score corresponding to each I.E. score. Our table would begin as shown in Table 21.

TABLE 21

SHOWING THE CORRESPONDENCE BETWEEN SCORES IN THE NATIONAL INTELLIGENCE TEST AND SCORES IN THE INTERMEDIATE EXAMINATION, THE PAIRS OF CORRESPONDING SCORES BEING DERIVED FROM THE LINE OF RELATION FOUND IN FIGURE 32

I.E.	N.I.T.	I.E.	N.I.T.	I.E.	N.I.T.	I.E.	N.I.T.	I.E.	N.I.T.
1	42	16	68						
2	43	17	etc.						
3	45	18							
4	47	19							
5	49	20							
		etc.							
6	50								
7	52								
8	54								
9	56								
10	58								
11	60								
12	61								
13	63								
14	65								
15	67								

EXERCISE 29. Complete Table 21 to show the correspondence as far as the I.E. scores of 73.

EXERCISE 30. What scores in the National Intelligence Test correspond to scores of 8, 16, 24, 32, and 40, respectively, in the Intermediate Examination?

EXERCISE 31. Show how to begin a table for convenient use in converting any N.I.T. score into the corresponding I.E. score.

**Alternative methods.** There are various other ways of finding a line of relation between two tests, which are not so accurate in that they take no account of any differences that may exist between the forms of the distributions of scores in the two tests, but are somewhat simpler to obtain.

One such method is simply to plot the pair of scores in the two tests having percentile ranks of 75 and the pair having

percentile ranks of 25 — in other words merely the upper- and lower-quartile scores in the two tests — and to draw a straight line through these points. This, of course, assumes that the line of relation is straight throughout, whereas we have seen that this may not be the case. To that extent the method is open to objection.

Another method assumes that the mean of one distribution corresponds to the mean of the other, that the line of relation is straight throughout, and that the units of the two variables<sup>1</sup> are so related that the number of units in the mean deviation of one distribution corresponds to the number of units in the mean deviation of the other. The line of relation is then drawn as shown at the left in Figure 35.

Still another method may be used which is the same as that just described, except that the standard deviation is used instead of the mean deviation. In that case the line of relation is drawn as shown at the right in Figure 35.

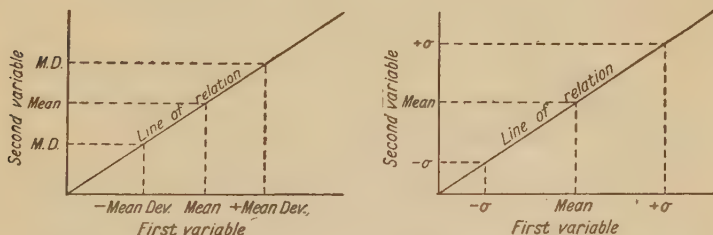


FIG. 35. Showing two methods of finding a line of relation between two variables, assuming the line to be straight.

### QUESTION

It has been proposed by one writer to compare scores in two tests by equating merely the total ranges of scores of a large group in the two tests. What do you think of the validity of that method compared with those suggested in this chapter?

<sup>1</sup> Scores, ratings, or any other measures.

## CHAPTER ELEVEN

### AVERAGING MEASUREMENTS OF A PUPIL

**How to average scores in two tests.** It is often desirable to average the scores of a pupil in two or more tests to find a single measure of his ability. Thus, we might want to combine the scores of our pupils in the National Intelligence Test and the Intermediate Examination in order to obtain a single measure of mental ability.

It is obvious that if we simply find the average of the scores of each pupil in the two tests, just as they stand, this average would not be comparable with scores in either test and, therefore, not comparable with norms furnished for either test. For example, a score of 46 in the Intermediate Examination may be seen in Figure 34 to correspond to a score of 120 in the National Intelligence Test. If a child made just these scores, the average of his scores would be 73, which is not comparable either with the 46 or with the 120.

To average scores in two tests just as they are is like averaging inches and centimeters. They should be converted to the same terms before finding the average.

When tables are provided for converting scores in both tests into terms of mental age, one method of bringing scores to the same terms is to convert each into a mental age.<sup>1</sup> The average mental age of the child by the two tests may then be found, and this will be directly comparable with either mental age alone. For reasons that will be brought out later, however, this method has its limitations, and at best it is indirect.

**Disadvantages of averaging percentile ranks.** The method has been used of averaging the percentile ranks of a pupil in two tests, and if the two percentile ranks are between

<sup>1</sup> See Chapter XII for the meaning of the term *mental age*.

25 and 75 this method is permissible. As a universal method, however, it is statistically objectionable because of the inequality of the values of units of percentile rank in different parts of the scale. The reason may be made clear by a simple illustration.

If you will consult Figure 30, you will see that a 75-percentile score in the hypothetical typical distribution is 72 points and a 99-percentile score is 97 points. Now the average of scores of 72 and 97 is  $84\frac{1}{2}$  and the percentile rank of this average score of 84 is about 93, whereas the average of the percentile ranks of 75 and 99 is only 87 (6 points too low).

This sort of discrepancy occurs only in the extremes, where there is a good deal of curvature to the percentile curve. In the middle half, where the curve is nearly straight, the method is reasonably valid and in general not a great deal of error will be introduced provided the percentile ranks to be averaged are not too far apart. However, there is a method that is universally applicable and valid. This will be described.

**The method of transmuting scores.** The best way to convert scores to the same terms is to transmute the scores of the pupils in one test into terms of the other. Thus, let us say that a pupil has made a score of 42 in the Intermediate Examination and a score of 120 in the National Intelligence Test. We can either transmute the score of 42 in the Intermediate Examination into terms of the National Intelligence Test (the National Intelligence Test equivalent being 114) and average this with 120, which will give us 117 as the average score of the pupil in terms of National Intelligence Test scores, or we can transmute the National Intelligence Test score of 120 into terms of the Intermediate Examination (the equivalent score being 46) and average this with the score of 42 made by the pupil in

the Intermediate Examination. In the latter case the pupil's average score is 44, in terms of scores in the Intermediate Examination, which you will see by your completed Table 21 to be the equivalent of 117 in the National Intelligence Test. In either case our average scores are all directly comparable with any norms that may be provided for one or the other of the two tests. Indeed, it will be seen that by converting both ways, the average scores can be found both in terms of the National Intelligence Test and in terms of the Intermediate Examination, so that comparison may be made with the norms of both tests.

There is another reason why the scores of a pupil in two tests should not be averaged directly when the variability of scores in one test is greater than the variability of scores in the other test, even though we are not interested in comparing scores with norms. The reason is that in such a case the finding of an average results in the giving of greater weight to the score of the test that has the greater variability of scores. This fact is illustrated in the following example:

Let us suppose that one pupil made a score of 54 in the Intermediate Examination and a score of 170 in the National Intelligence Test; another pupil made a score of 70 in the Intermediate Examination and a score of 132 in the National Intelligence Test. (See Figure 36.) Now an I.E. score

of 54 corresponds to an N.I.T. score of 132, according to the line of relation in Figure 34, and an I.E. score of 70 corresponds to an N.I.T. score of 170. It will be seen that each pupil has made the equivalent of scores of 54 and 70 in the Intermediate Examination or of 132 and 170 in the National Intelligence Test, and the average

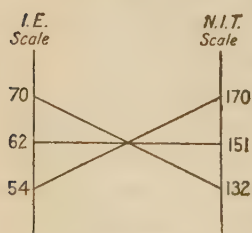


FIG. 36.

scores of the two pupils should be the same, no matter whether found in terms of the one test or of the other. That is, if the average scores of the two pupils are found by transmuting their National Intelligence Test scores into terms of the Intermediate Examination, these averages will both be 62, and if the averages are found by transmuting the Intermediate Examination scores into terms of the National Intelligence Test, the average scores of both pupils will be 151, which, of course, corresponds approximately to a score of 62 in the Intermediate Examination.

However, if we should average the scores directly, — i.e., without transmuting, — we should find that the average score of the first pupil is 112 and the average score of the second pupil is 101, a difference of 11 points. As has been stated above, the difference between these averages is due to the fact that the scores in the National Intelligence Test are more heavily weighted in averaging than the scores in the Intermediate Examination, not because of their being larger scores, but because of the variability being greater or, in other words, because the difference between 132 and 170 is greater than the difference between 54 and 70.

Of course, it will seldom happen that a pupil makes scores in two tests differing as widely as 54 and 70 in the Intermediate Examination or 132 and 170 in the National Intelligence Test, but this exaggerated case illustrates the principle involved in all cases; namely, when the variability of scores in one test is greater than the variability of scores in another, averaging scores in the two tests directly results in giving greater weight to the score in the test having the greater variability of scores.

You will be safe in following the general principle that whenever scores in two tests, which are not directly comparable in themselves, are to be averaged, the scores of one test should be transmuted into terms of the other test before

averaging. When the scores in the two tests are directly comparable, as for example scores in the two forms of the same test given under such conditions that practice effect, growth in ability between tests, etc., may be considered negligible, the scores may be averaged directly. It is always safe, however, to determine the variability of the scores in each test by means of a percentile graph from which may be determined also whether the scores in one test tend to run higher than those in the other test. If the scores in one test are either more variable or, on the average, higher than those in the other test, the transmuting should be done before averaging.

**The need for a universal standard scale.** Considering that the scores of no two tests are directly comparable and that, unless one takes the trouble to find the correspondence in a manner such as the one explained above, it is not possible ordinarily to compare the scores of pupils in two different tests, the need has been felt for some common terms in which the scores of the various tests could be expressed — a sort of universal language into which all scores could be translated.

No entirely satisfactory standard scale has been devised to date, but one such scale — the *T*-score scale — will be described to show the trend.

***T* scores.** Dr. McCall has proposed a scale of units called *T* scores<sup>1</sup> into which the scores of any test may be transmuted in order to be compared with the scores of any other test the scores of which are similarly transmuted.

The *T*-score scale is based upon the hypothetical distribution of ability of a standard group of individuals. This standard group consists of all children between the ages of 12 and 13 years. For practical purposes any large group

<sup>1</sup> See William A. McCall, *How to Measure in Education*; The Macmillan Company.

of unselected pupils whose age last birthday was 12 years is considered a standard group.

We may think of a  $T$  score as a score in an imaginary test so constructed that the distribution of scores of our standard group of 12-year-olds is a normal distribution, with a mean score of 50 and a standard deviation of 10 points. Converting the scores of any test into  $T$  scores, therefore, is merely to convert them into terms of this imaginary test.

Now in a normal distribution the percentile rank of an individual making a score of Mean +  $1\sigma$  is 84, the percentile rank of an individual making a score of Mean +  $2\sigma$  is 97.7, etc., as shown in Table 22.

TABLE 22

SHOWING THE PERCENTILE RANKS OF VARIOUS SCORES IN A NORMAL DISTRIBUTION IN TERMS OF THE STANDARD DEVIATION OF THE DISTRIBUTION

SCORE	P.R.	T.S.	SCORE	P.R.	T.S.	SCORE	P.R.	T.S.
$M + 3\frac{1}{2}\sigma$	99.98	85	$M + \sigma$	84	60	$M - 1\frac{1}{2}\sigma$	7	35
$M + 3\sigma$	99.87	80	$M + \frac{1}{2}\sigma$	69	55	$M - 2\sigma$	2.3	30
$M + 2\frac{1}{2}\sigma$	99.4	75	Mean	50	50	$M - 2\frac{1}{2}\sigma$	0.6	25
$M + 2\sigma$	97.7	70	$M - \frac{1}{2}\sigma$	31	45	$M - 3\sigma$	0.13	20
$M + 1\frac{1}{2}\sigma$	93	65	$M - \sigma$	16	40	$M - 3\frac{1}{2}\sigma$	0.02	15

Now, in making up the  $T$  scale to correspond to the scores of any test, the mean score of our standard group is assigned a  $T$ -score value of 50, the score having a percentile rank of 84 ( $M + \sigma$  in a normal distribution) is assigned a  $T$ -score value of 60, the score having a percentile rank of 97.7 ( $M + 2\sigma$  in a normal distribution) is assigned a  $T$ -score value of 70, etc., as shown in Table 22. The  $T$ -score values, therefore, are assigned on the basis of percentile rank in such a way that if the distribution of scores in the test is normal, each unit of  $T$  score represents  $\frac{1}{10}$  of the standard deviation of the distribution. When a table of  $T$ -score equivalents of scores in any test has once been made up, scores may be

converted into *T*-score equivalents, of course, merely by referring to the table.

**Disadvantages of the *T*-score method.** The *T* scale as a universal scale has certain disadvantages which are fully appreciated by Dr. McCall.

The major weakness of the *T*-scale method is that it is applicable only to tests that are suitable for testing 12-year children. The scores of a kindergarten test, for example, which is so easy that practically all 12-year children would get perfect scores cannot be converted into *T* scores, and similarly the scores of a test for high school students which is too hard for 12-year-olds cannot be converted into *T* scores.

But even in the case of tests suitable for 12-year children the *T*-score values obtained in the extremes of the distribution are very unreliable, since, as you will readily see, the slight difference in percentile rank of a certain score between 99.89 and 99.98 means a difference of 5 points in *T* score — as much as the difference between percentile ranks of 50 and 69.

If any universal scale such as the *T* scale is to be practical, therefore, it must be based on distributions of more than one age group. Thus we might use the 12-year age group for *T* scores only between 40 and 60 and use, say, an 8-year group for *T* scores below 40 and a 16-year group for *T* scores above 60. The three sections of the scale, of course, would have to be properly integrated by extended research as to the typical overlapping of ability of 8-, 12-, and 16-year age groups. This research would constitute a good subject for a master's thesis.

#### HOW TO AVERAGE SCORES AND TEACHER'S MARKS

It is often desirable to base promotions on the basis of the average of a pupil's score in an achievement test and his scholarship mark assigned by the teacher as an estimate of the quality of school work he has done throughout the

year or term. Thus, a suitable basis for promotion would be the average between the pupil's score in the Achievement Test which constitutes Part I of the Otis Classification Test,<sup>1</sup> and the mark which the teacher has assigned as a measure of the pupil's scholarship for the term.

TABLE 23

SHOWING THE SCORES OF 43 SIXTH-GRADE PUPILS IN THE ACHIEVEMENT TEST AND THEIR TEACHER'S MARKS IN SCHOOL WORK

PUPIL	TEA. MARK	SCORE A.T.	PUPIL	TEA. MARK	SCORE A.T.	PUPIL	TEA. MARK	SCORE A.T.
1	78	75	16	91	54	31	64	42
2	91	83	17	75	66	32	86	50
3	83	49	18	88	67	33	87	72
4	63	45	19	77	71	34	96	57
5	76	61	20	61	48	35	68	41
6	79	56	21	73	41	36	79	58
7	79	73	22	65	64	37	78	54
8	84	53	23	82	64	38	80	36
9	71	63	24	88	76	39	77	64
10	62	47	25	77	45	40	71	59
11	80	45	26	78	33	41	78	81
12	77	63	27	76	63	42	80	47
13	95	73	28	74	49	43	91	77
14	93	58	29	73	49			
15	93	91	30	70	41			

In Table 23 are shown the scores of our 43 6A pupils in the Achievement Test, Form A, and their teacher's mark in school work. Let us suppose that we wish to find the average of the score and teacher's mark of each pupil.

As explained above, it is advisable in such a case to represent the distributions of scores and teacher's marks by percentile curves, in order to determine whether one distribution has a greater variability than the other and what tendency there may be for the teacher's marks to run higher or lower than the scores.

<sup>1</sup> Published by World Book Company, Yonkers-on-Hudson, New York,

UNIVERSAL PERCENTILE GRAPH

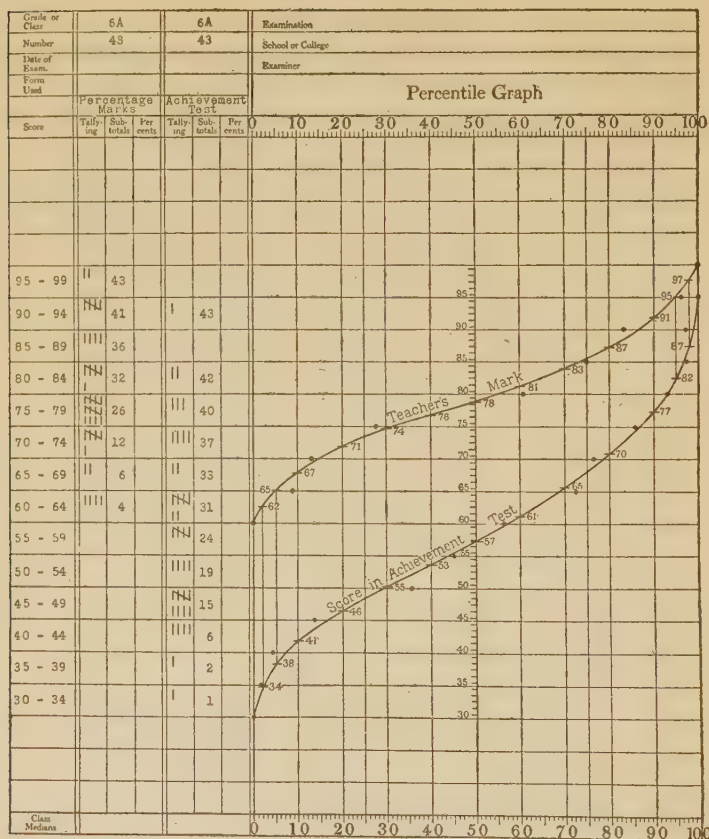


FIG. 37. The percentile curves representing the distributions of scores in the Achievement Test and of teacher's marks of 43 pupils in Grade 6A, showing the method of finding the scores corresponding to various selected teacher's marks.

In Figure 37 are shown the percentile curves that represent the distributions of scores and teacher's marks of the 6A pupils. It is instantly apparent from these curves that

the teacher's marks tend to run much higher than the scores. It may be seen at a glance also that the variability of teacher's marks is less than the variability of scores. This is shown by the fact that the curve representing the distribution of teacher's marks is not so steep as the curve representing the scores.

TABLE 24

SHOWING THE SCORE IN THE ACHIEVEMENT TEST AND TEACHER'S MARK CORRESPONDING TO EACH OF VARIOUS SELECTED PERCENTILES OF GRADE 6A, OBTAINED FROM THE PERCENTILE CURVES IN FIGURE 37 FOR USE IN DRAWING THE LINE OF RELATION BETWEEN THESE VARIABLES

PERCENTILE	SCORE	TEACHER'S MARK
98	87	97
95	82	95
90	77	91
80	70	87
70	65	83
60	61	81
50	57	78
40	53	76
30	50	74
20	46	71
10	41	67
5	38	65
2	34	62

For these reasons it is obvious that scores and teacher's marks are not directly comparable and that, in order to obtain a valid and interpretable average between them, it is necessary either to transmute the scores into terms of teacher's marks or teacher's marks into terms of scores. To do this we must make up a table, as before, showing the correspondence between scores and teacher's marks. This correspondence can be found in exactly the same way as we found the correspondence between scores in the Intermediate Examination and scores in the National Intelli-

gence Test. That is, we can plot on cross-section paper a point representing the 2-percentile score and the 2-percentile teacher's mark; another point representing the 5-percentile score and mark; another representing the 10-percentile score and mark, etc., as before. If we arrange our pairs of values in a table as before, preparatory to plotting the points, our table will appear as shown on previous page.

Having drawn up a table of corresponding scores and teacher's marks, we are ready to plot these in a chart and draw the line of relation between scores and teacher's marks. This has been done in Figure 38.



FIG. 38. Showing the line of relation between scores in the Achievement Test and teacher's marks in Grade 6A, located by plotting points representing pairs of percentile values obtained from Figure 37.

Note that there are two purposes in drawing this line of relation. One is to smooth out the irregularities that occur in our attempt to equate single scores and teacher's marks having the same percentile rank, and the other is to enable us to *interpolate* — that is, to fill in the intervening values between those taken from the percentile curves.

In drawing the line of relation in Figure 38, it was assumed that the points lay so nearly in a straight line that for practical purposes it might be considered that the relation between scores and teacher's marks was best represented by a straight line.

TABLE 25

SHOWING THE CORRESPONDENCE BETWEEN SCORES IN THE ACHIEVEMENT TEST AND TEACHER'S MARKS IN SCHOOL WORK FOR 43 SIXTH-GRADE PUPILS

TEA. MARKS	SCORE	TEA. MARKS	SCORE	TEA. MARKS	SCORE	TEA. MARKS	SCORE
60	30	70	45	80	60	90	75
61	31	71	46	81	61	91	76
62	33	72	48	82	63	92	78
63	34	73	49	83	64	93	79
64	36	74	51	84	66	94	81
65	37	75	52	85	67	95	82
66	39	76	54	86	69	96	84
67	40	77	55	87	70	97	85
68	42	78	57	88	72	98	87
69	43	79	58	89	73	99	88

We are now ready to make up a table of correspondence between teacher's marks and scores in the Achievement Test from this line of relation. If we plan to convert teacher's marks into terms of scores, we should begin our table by writing in consecutive order the various teacher's marks, beginning with the lowest teacher's mark that we might have, and then write opposite each teacher's mark the score

that corresponds to it, according to the line of relation. Thus, it will be seen that, according to the line of relation, the score corresponding to a teacher's mark of 60 is 30; the score corresponding to a teacher's mark of 61 is 31; etc., as shown in Table 25.

We are now ready to find the average of any pupil's score and teacher's mark. Let us take Pupil No. 1 in Table 23, for example, whose score was 75 and whose teacher's mark was 78. It will be seen by Table 25 that a teacher's mark of 78 corresponds to a score of 57. This shows that the pupil's mark was in reality very much lower than his score, relatively speaking. The average of his actual score of 75 and the score of 57 corresponding to his teacher's mark is 66. Pupil No. 3 had a teacher's mark of 83, which, according to Table 25, corresponds to a score of 64, showing that this pupil's mark was relatively much higher than his score, which was 49. His average is the average between 64 and 49, which is  $56\frac{1}{2}$ .

Having found the averages of the scores and teacher's marks of all the pupils in the class, we may then arrange these in order of magnitude and proceed to use them as the basis for promotion. Undoubtedly promotions based on averages so found would be more just to the pupils than if based on either teacher's marks or scores in the Achievement Test alone.

**EXERCISE 32.** Find the average score and teacher's mark in terms of score, of each of Pupils 1 to 10, and write their numbers in the order of their averages, beginning with the highest.

You will realize, of course, that it is possible to average the scores in any number of tests together with any number of teacher's marks by this same method; that is, by transmuting all the scores and teacher's marks into terms of some single score and then finding the average of these transmuted

values for each pupil. Thus, promotion might be based upon average of teacher's marks and scores in an achievement test and a mental-ability test.

If it is desired to give equal weight to each of these three measures in finding the average, the three measures should be reduced to common terms; as, for example, by transmuting scores in the mental-ability test and teacher's marks into terms of scores in the achievement test and finding the average of these three measures for each pupil.

If, on the other hand, it is desired to give as much weight to the teacher's mark as to the scores in the two tests together, this may be done as follows: Transmute the three measures into the same terms, as explained previously, and then average these values, counting the transmuted teacher's marks twice and dividing by 4 instead of 3. For example, if a pupil's achievement score is 50, his transmuted mental-ability score is 54, and his transmuted teacher's mark is 60, his weighted average would be found by taking the average of 50, 54, 60, and 60, which is 56.

Any combination of test scores and teacher's marks may be averaged, giving any desired weight to any measure of ability. The weight that is given any measure is in proportion to the number of times it is counted in finding the average.

You may be interested to note that when two numbers (say 60 and 69) are averaged giving double weight to the second, the weighted average (66) is just  $\frac{2}{3}$  of the way from the first to the second. If the second number is given three times the weight of the first, the weighted average is  $\frac{3}{4}$  of the way from the first to the second, etc.

**EXERCISE 33.** Find the weighted average of the scores and teacher's marks of Pupils 1 to 10 in Table 23 in terms of scores, giving to the score twice the weight that is given to the teacher's mark. (Add the score twice and divide by 3.) Rank these pupils by number in the order of their weighted averages, beginning with the highest.

## CHAPTER TWELVE

### THE GROWTH OF MENTAL ABILITY

**A growth chart.** You have now had some practice in plotting points on cross-section paper to represent pairs of values. You will find this method a serviceable one for many purposes. For example, it may be used in a very helpful way to indicate the manner of growth of a boy or a girl from year to year in any measurement or trait, such as height, weight, or mental ability.

Thus let us consider first the more concrete case of the growth of a hypothetical boy, Frank, in height. Suppose that we had measured Frank's height on his birthday every year from birth to the age of 20 years and found his height at these various ages to be as shown in Table 26.<sup>1</sup>

TABLE 26

SHOWING THE HEIGHT OF A HYPOTHETICAL BOY, FRANK, AT EACH AGE FROM BIRTH TO 20 YEARS

AGE IN YEARS	HEIGHT IN INCHES	AGE IN YEARS	HEIGHT IN INCHES	AGE IN YEARS	HEIGHT IN INCHES
0	22	7	$47\frac{3}{4}$	14	63
1	29	8	50	15	$65\frac{1}{4}$
2	$33\frac{1}{2}$	9	52	16	$67\frac{1}{4}$
3	$37\frac{1}{4}$	10	54	17	$68\frac{1}{2}$
4	$40\frac{1}{4}$	11	56	18	$69\frac{1}{2}$
5	43	12	58	19	70
6	$45\frac{1}{2}$	13	$60\frac{1}{2}$	20	70

**How to draw a growth chart.** If we take a sheet of cross-section paper and lay out along the bottom an age scale, as shown in Figure 39, and on the side a scale of heights in

<sup>1</sup> These figures show the normal or median heights for the various ages as given in the Baldwin-Wood Age-Height-Weight Table.

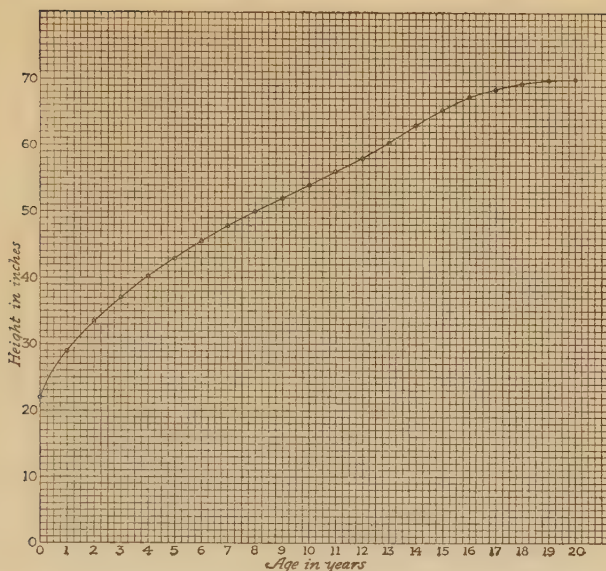


FIG. 39. Showing the growth in height of a hypothetical boy, Frank, from birth to 20 years of age.

inches, then we may represent Frank's height at birth (22 inches) by a point above zero on the age scale and having a height above the base line equal to 22 units by the height scale. Frank's height at the age of one year (29 inches) may be represented by a point above 1 on the age scale and having a height above the base line of 29 units by the height scale. And so on.

If we plot a point in our graph to represent Frank's height at each age up to 20 years, the points will appear as shown in Figure 39. In this figure a smooth curve is drawn through the points, and presumably if we had plotted any other points, — such, for example, as those representing Frank's height at the ages of one-half year, and one and one-half years, etc., — these would fall on the same curve. Indeed,

it would be possible to draw the curve almost as accurately with only every other point. This is because the measurements may be made accurately and because a boy grows in a fairly regular or steady manner.

**Interpretation of a growth chart.** By means of the curve it is possible to tell with considerable accuracy just how tall Frank was at any age, such as 10 years, 3 months. (Each small unit of horizontal distance represents 3 months in age.) According to the graph in Figure 39, Frank was at age 10 years 3 months,  $54\frac{1}{2}$  inches in height.

The curve shows a number of aspects of Frank's growth that are not evident from the figures given in Table 26. Thus, he may be seen to have grown comparatively rapidly during the first year of life and then to have grown less rapidly until the age of about 7 years, at which time the rate of growth became comparatively constant up to the age of about 12 years, when there was a spurt for two or three years. Then the rate gradually diminished to zero. Very little growth took place after the age of 17 years. The steeper the curve, the more rapid the growth. When the curve has become level, growth has ceased.

**Comparison of growth curves.** It is a matter of common observation that children differ appreciably in their heights at the same age and reach adult stature at different heights and at different ages as well.

It is possible, of course, to draw the curve of growth of two or more individuals on the same chart. Thus Figure 40 shows the same curve of growth as is shown in Figure 39, accompanied by a second curve representing the growth of another boy, George.

How tall was George at birth? <sup>1</sup> at the age of 2 years? at the age of 5 years? Each vertical unit represents how many inches in height? <sup>2</sup> At the age of 10 years, was George as

<sup>1</sup> 20 inches.

<sup>2</sup> 1 inch.

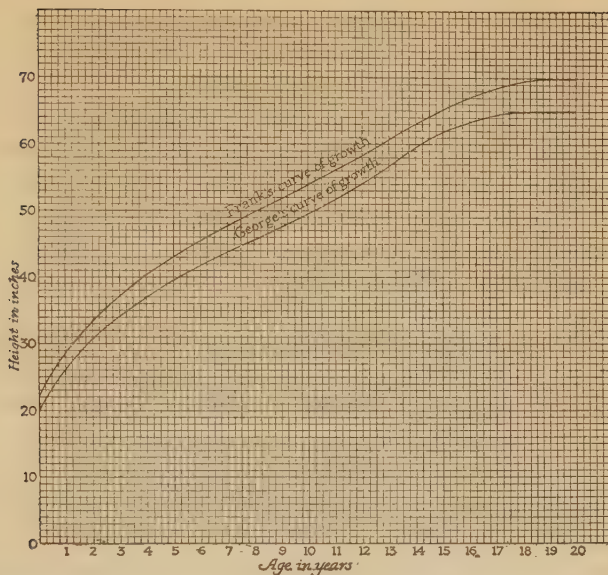


FIG. 40. Showing the relation between the growth of two boys who grew at different rates and attained adult stature at different ages and heights.

tall as Frank? What was the difference in their heights at that age?<sup>1</sup> At what age did George reach the height Frank had at the age of 6 years, 3 months?<sup>2</sup> At about what age was George growing most rapidly?<sup>3</sup> Which boy was growing the more rapidly at the age of 13 years? of 16 years? At about what age did George reach adult stature? At what age did Frank reach this height? By how many inches did Frank's adult height exceed George's?

**Growth in mental ability.** It is possible to represent a child's growth in mental ability by a curve similar to those we have used to represent growth in height. Of course we cannot measure mental ability as accurately as we can

<sup>1</sup>  $4\frac{1}{2}$  inches.

<sup>2</sup> 8 years, 3 months.

<sup>3</sup> 13 years.

measure height. How do we measure mental ability? Indeed, what is mental ability?

**Definition of mental ability.** We cannot, of course, digress here to the extent of discussing the psychological considerations underlying a definition of mental ability. But we can make a definite statement about it. We can say that mental ability is that quality of the mind which enables a person to apply the knowledge he possesses to the solution of new problems. It is a quality which possesses magnitude in the sense that a child as he grows older shows the ability to apply his knowledge to the solution of more and more "difficult" problems. (A *difficult* problem or situation is one that fewer individuals of a given age can solve or cope with successfully than an *easy* one.) Thus, if we draw up a series of questions and problems (let us call them *items*) ranging from easy to difficult, we find that a child as he grows older can answer or solve more and more of these items. This fact gives us a convenient method of measuring mental ability. We can give the pupils of a class the Intermediate Examination and count the items each pupil got right and call that his score. Then in so far as the pupils have had the *same opportunity to acquire knowledge* (for, indeed, the amount of knowledge acquired by a pupil up to a given age depends to a large extent on the rate of his growth in mental ability), to that extent a pupil's score in a mental-ability test is one measure of mental ability.

**Validity of mental-ability test scores.** Needless to say, if two pupils have not had the same opportunity and motivation to acquire knowledge, due, in one case, to sickness, poor instruction, or similar cause, then the scores of these two pupils cannot show their true relative degrees of mental ability. For that reason we must remember that *a mental-ability test does not measure mental ability directly*. In most cases, however, pupils of the same age in the same school

have had approximately the same opportunity to acquire knowledge and to that extent the mental-ability test does measure their relative degrees of mental ability.

Let us consider this point in the light of an analogy. Measurement of the weight of water that various vessels hold is an indirect measure of the capacities of the vessels. If the vessels were not all full, the weights of water might fail to show their relative capacities; but if they were all full, then the weights of water would be perfectly valid measures of the relative capacities of the vessels. To be sure, if the vessels were all just  $\frac{3}{4}$  or  $\frac{7}{8}$  full, in this case also the weights of water would show the true relative capacities of the vessels.

Similarly, children to be adequately measured in mental ability by a "mental-ability test" must be "full" of knowledge up to their capacities at the time, or at least "filled" to the same relative degree. This means, as we said before, that two pupils must have had educational opportunities (instruction and motivation) enabling them to acquire knowledge up to the same proportion of their capacities in order to be properly compared in mental ability by their scores.

In absence of any adequate information as to the relative educational opportunities of the pupils of a group, however, we can only fall back on their *scores* as measures of their relative degrees of mental ability.

**Units of measurement of mental ability.** Are the units of measurement equal in the sense that inches are equal? We usually think of 10 points of score as equal to 10 points of score, no matter whether it is the difference between scores of 20 and 30 or the difference between scores of 65 and 75. As a matter of fact, however, we know that 10 points in one portion of the scale of scores of a test may not at all equal 10 points in another portion of the scale. We cannot prove this with one test alone, but if we refer to

Figure 34 we see that 10 points in I.E. score between 40 and 50 correspond to 14 points in N.I.T. score (111 to 125), whereas 10 points in I.E. score between 60 and 70 correspond to 23 points in N.I.T. score (146 to 169). Obviously, the values of the units in either one or the other of these scales must vary from one part of the scale to another; and this may be, and probably is, true of both scales.

**A curve of growth in mental ability.** In spite of the fact that we know that the true values of units of the scale of scores in a test probably vary somewhat, a pupil's score in a mental-ability test is a fairly satisfactory measure of his mental ability. Let us draw Frank's curve of growth in mental ability as measured in terms of score in the Intermediate Examination. We shall suppose Frank to have been tested on his birthday each year from the age of 8 years to the age of 18 years, and that his scores were as shown in Table 27.

TABLE 27

SHOWING FRANK'S SCORES IN THE INTERMEDIATE EXAMINATION FROM  
8 TO 18 YEARS

Age	8	9	10	11	12	13	14	15	16	17	18
Score	8	15	24	26	34	40	50	54	56	59	59

**Drawing the curve of growth.** Our next step, of course, is to plot points on cross-section paper representing these pairs of values. This has been done in Figure 41 and a smooth curve drawn "through" the points. Why a smooth curve? Why not just join the points with straight lines?

Is it likely that a child's true mental ability would grow in the manner that would be indicated by joining the points with straight lines? Why?

**Regular growth with fluctuations of measurement.** If we assume that Frank's mental ability increased in a regular

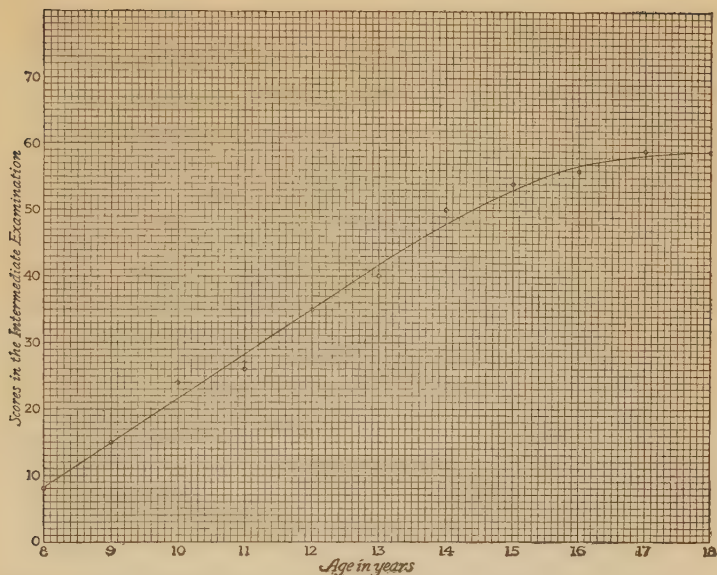


FIG. 41. A hypothetical curve of growth in mental ability as measured by the Intermediate Examination.

manner and that the most probable true curve of growth is that which is drawn in Figure 41, how shall we account for the fluctuation of points above and below the line?

There are, of course, many reasons why a pupil's score on any particular day may not indicate his true ability relative to his other scores and to the scores of the other members of the class. You will remember from a previous discussion that fluctuating educational opportunity (instruction and stimulation) affect scores. Also mental and physical conditions (i.e., nervousness or fatigue) influence scores; etc. Is it not reasonable to assume, therefore, that Frank's score at age 10 was too high and his score at age 11 too low, and that the line strikes a sort of average between these two?

Indeed, we may assume the curve in Figure 41 to be in the nature of an average between all the scores that are plotted. Each one above the line is balanced by one below the line, very much as each measure above an average is balanced by one below it. The line bears the same relation to the increasing scores that the "dead level" at *B* and *D* in Figure 1 bore to the heights of liquid at *A* and *C*.

You may be inclined to ask how a pupil's score could be too high. Can a pupil do better than his best? Should we not consider his best score as his true one? Perhaps so. In that sense there should be no points above the line. However, if we could give Frank ten forms of the Intermediate Examination on succeeding days, the average or median score would be a more stable and representative measure of his everyday ability than his highest score. The scores above the line, therefore, merely represent performances made under conditions more favorable than the average conditions under which tests are taken.

**Interpreting the curve of growth.** Assuming our curve to represent best Frank's growth of mental ability, what are the outstanding characteristics of this growth? It proceeded at a uniform rate from age 8 to about age 14, when a slowing in growth began and continued to the age of about 18, at which point growth apparently stopped. At that age he is said to have reached *mental maturity*.

**Individual differences.** Children differ even more in their growth in mental ability than in their growth in height. Thus Figure 42 shows how three boys might differ in their growth in mental ability.

You will note that Harold showed greater mental ability at 8 than Frank and retained this superiority throughout his growth, reaching mental maturity at a higher point in the scale; and similarly, George, who showed less mental ability than Frank at 8 years, retained his inferiority throughout his growth, reaching mental maturity at a lower point in the scale.

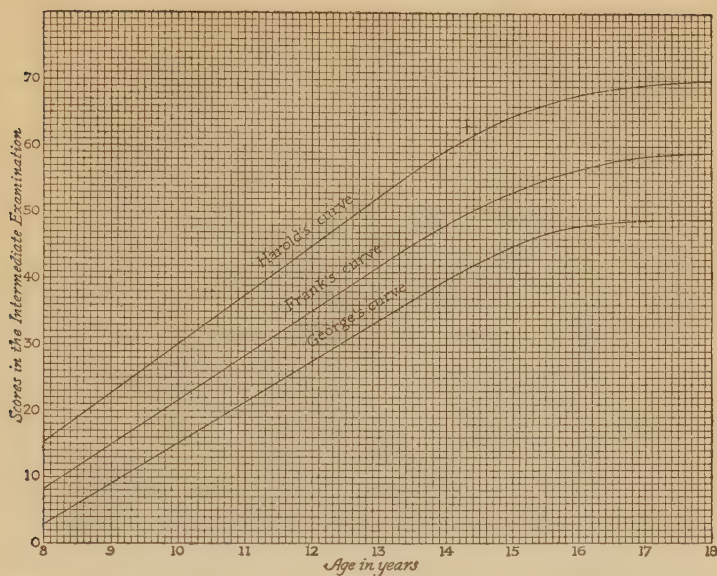


FIG. 42. Curves of growth in mental ability of three hypothetical boys, showing individual differences.

**Bright, normal, and dull children.** Seeing the large differences in the scores of pupils of the same age, the question naturally arises as to the cause of these differences. Except for a certain degree of variability due to the errors of measurement, we do not know the cause any more than we know why some children are tall for their ages and some are short. We can merely say that some children are *bright*, some are *dull*, and some are just *normal*.

What constitutes a normal child? A child of ten years is said to be exactly normal if his score is the median of the scores of 10-year children.<sup>1</sup>

<sup>1</sup> We cannot measure all 10-year children, of course; so we measure as large a sampling of 10-year children as we can get. By 10-year children is meant theoretically children who are just 10 years old exactly. In practice we find the median score of pupils from  $9\frac{1}{2}$  to  $10\frac{1}{2}$  years.

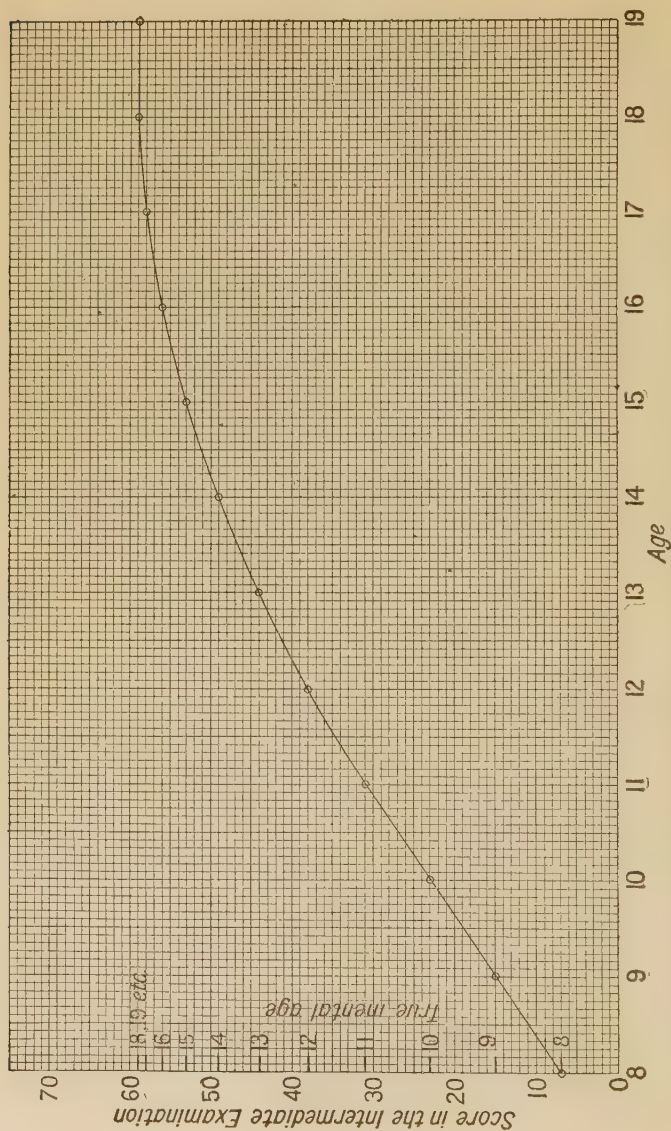


FIG. 43. The normal curve of growth in mental ability as measured by the Intermediate Examination.

Thus, the median score of 10-year children in the Intermediate Examination is 23 points. A 10-year pupil who makes a score of 23 points in the Intermediate Examination, therefore, is said to be just normal.

**Mental age.** A pupil of any age making a score of 23, therefore, has the mental ability that is just normal for the age of 10 years. To all intents and purposes he is just 10 years old mentally. For convenience, then, we say that such a pupil has a *mental age* of 10 years. The expression "a mental age of 10 years" has a double usefulness. It is a common measure to which scores of other tests can be converted for ease in comparison (it is merely a matter of finding the median score of 10-year children in the tests in question), and it is "self-interpretive," as we have said. We can all easily understand and remember what it means to say that a pupil has a mental age of 10 years.

**Norms.** The median score of 11-year children in the Intermediate Examination is 31. A pupil of any age making a score of 31, therefore, is said to have a mental age of 11 years. The score of 31 is said to be the *norm* for the age of 11 years.

Figure 43 shows the normal curve of growth in mental ability as measured by the Intermediate Examination. Note that the curve passes through a point above 11 on the age scale and opposite 31 on the score scale. Similarly, as shown by the other points through which the curve passes, the score of 38 is the norm for the age of 12 years, the norm for the age of 13 years is 44, etc.

**EXERCISE 34.** What is the norm for the age of 14 years? The score of 53 is the norm for what age? What is the norm for the age of  $9\frac{1}{2}$  years? 10 years, 3 months? 11 years, 5 months? 16 years? 17 years? 18 years? 19 years? 20 years? (Obviously the norm for all ages above 18 years is the same as the norm for 18 years and is 59.) The score of 21 is the norm for what age? For what age is the score of 37 the norm? For what age is 46 the norm? What, then,

is the mental age of a pupil making a score of 46? What is the mental age of a pupil making a score of 47? of 51? of 57? of 58? of 59? of 60?!

**The limit of true mental ages.** A pupil making a score of 58 has a mental age of 17 years. This is shown both by the curve and by the scale at the left in Figure 43, which is derived from the curve. A pupil making a score of 59 has a mental age of 18. A pupil making a score of 60 has a mental ability above that normal for any age. Obviously, therefore, no score above 59 in the Intermediate Examination can be expressed in terms of mental age, if by mental age we mean the age of the exactly normal individual making that score, because the exactly normal individual never makes a score above 59, having reached mental maturity at that point in the scale.

**Binet mental ages.** The term "mental age" was used first in connection with the Binet-Simon Tests, which antedated all group tests. Scores in this test <sup>1</sup> are not ordinarily given in terms of number of questions correctly answered, but are calculated in terms of years and months and are called *mental ages*.

The Binet-Simon Test as first devised did not test mental ability above that normal for adults, but it was revised and extended by Dr. Lewis M. Terman of Stanford University in what is known as the Stanford Revision of the Binet-Simon Test <sup>2</sup> so that it would measure mental ability much above that normal for adults. Now, in order to preserve the previous nomenclature and to be consistent, these degrees of mental ability above that normal for adults were still expressed in terms of years and months and called mental

<sup>1</sup> The Binet-Simon Tests, so called, constitute a single test, as we ordinarily think of a test nowadays; that is, a single list of questions arranged in a series. We shall therefore speak of the Binet-Simon Tests as a test.

<sup>2</sup> Published by Houghton Mifflin Company.

age — no distinction being made between these degrees of mental ability above the norm for adults and those degrees below the norm for adults.

The mental ages in the case of the Stanford Revision of the Binet-Simon Test extend from 3 years to 19 years, 6 months. The correspondence between various scores in the Intermediate Examination and Binet mental ages has been found in one study to be approximately as shown in Figure 44.

Observe this figure carefully. Note that up to the Binet mental age of 11 years Binet mental ages<sup>1</sup> correspond to the ages for which corresponding scores in the Intermediate Examination have been found normal. The Binet mental age,<sup>2</sup> according to this figure, that is normal for the age of 12 years is about 11 years, 11 months. The Binet mental age that is normal for the age of 13 years is about 12 years, 8 months, etc., and the Binet mental age that is normal for the age of 18 years is about 15 years, 5 months. It has been the custom to consider a Binet mental age of 16 years, 0 months as the norm for adults (individuals who have reached mental ma-



FIG. 44. Showing the relation between Binet mental ages and true mental ages in terms of scores in the Intermediate Examination.

<sup>1</sup> The Stanford Revision of the Binet-Simon Test is now frequently spoken of as the Binet Scale, and we shall use the shorter term Binet mental ages to refer to Stanford-Binet mental ages.

<sup>2</sup> Let us think of these Binet mental ages, as for example this mental age of 11 years, 11 months, as *scores* for the time being and not as mental ages in the sense we have been discussing on preceding pages.

turity), but in the light of army data psychologists are coming to believe the norm for adults to be below the Binet mental age of 16 years.

Under any circumstances a Binet mental age of 17 years represents a degree of mental ability much above the norm for the age of 17 years or for any age. A Binet mental age of 18 years, of course, represents, and is intended to represent, a degree of mental ability very much above the normal mental ability of 18-year-olds, as may be plainly seen from Figure 44. A Binet mental age of 19 years represents a degree of mental ability so high that probably only about 3 per cent of adults ever attain it.

**Fictitious mental ages.** Let it be perfectly understood, therefore, that *Binet mental ages in the upper ranges are NOT normal for the corresponding chronological ages. They are merely high scores that are expressed in terms of years and months.* To distinguish these mental ages from the true mental ages we have previously discussed, we may term these *fictitious mental ages*.

If a mental age represents a degree of mental ability that is normal for the corresponding chronological age, we may call it a *true mental age*. Binet mental ages in the lower ranges are true mental ages. If a mental age represents a degree of mental ability that is not normal for the corresponding chronological age, we should call it a fictitious mental age. Binet mental ages in the upper ranges, as we have said, are fictitious mental ages. The dividing line between the true and the fictitious mental ages in the Binet mental-age scale is not known exactly but is probably between 11 and 14 years.

The fictitious character of Binet mental ages in the upper ranges is perfectly understood by the authors of the various revisions of the Binet-Simon Test, but it is wholly unappreciated by a large number of persons who speak very familiarly of mental ages, norms, etc. Indeed, this misconception

probably has been the source of more confusion and misunderstanding as to the use of mental ages in interpreting scores, finding intelligence quotients, etc., than any other. When the special character of Binet mental ages is appreciated and it is realized that a true mental age of 18 years means one thing and a Binet mental age of 18 years means another thing, we need have no more confusion. It is necessary, however, to distinguish always between Binet mental ages and true mental ages when these are above 12 or 14. The best way is simply to refer always to mental ages that are obtained by or are in terms of the Binet-Simon Test as *Binet mental ages*. This cannot be too strongly emphasized.

#### QUESTIONS

1. The Manuals of Directions for various group mental tests give age norms. Let us suppose the norm for the age of 17 years is given as 129 points. Let us suppose a high school student has made a score of 129 in the test. May we say that he has a mental age of 17 years?

2. Let us suppose the student to be 16 years old. May we divide this mental age by his "chronological age" and find his IQ?<sup>1</sup>

3. If we did so, would his IQ so found be comparable with an IQ found by means of the Binet-Simon Test?

4. What must be done in order to get an IQ from the score and chronological age comparable with an IQ found by means of the Binet-Simon Test?

<sup>1</sup> See page 148 for meaning of IQ.

## CHAPTER THIRTEEN

### THE MEASUREMENT OF BRIGHTNESS

WE have seen how a pupil's degree of mental ability may be expressed either in terms of his score in a mental-ability test or in terms of "mental age." This, however, does not give us a complete account of the mental quality of the child. Thus, one pupil may have attained the mental age of 5 years (the mental status of a normal child of 5 years) when he was just 5 years old; another may have attained the mental age of 5 years when he was but 4 years old; and still another may not have attained the mental age of 5 years until he was 6 years old.

The intelligence quotient. Obviously there is a fundamental difference between these pupils even though all three have reached the same mental status. We call this a difference in *brightness*.

It is the same kind of difference that would exist between three boys — let us call them Harold, Frank, and George — who at the age of 10 years have mental ages of 11 years, 10 years, and 9 years, respectively.

Now in order to express the degree of brightness of each of these boys, we may divide his mental age by his chronological age,<sup>1</sup> and let the quotient express his degree of brightness. For example, Harold's mental age is 11 years and his chronological age is 10 years; hence his measure of brightness is  $\frac{11}{10}$  or 1.10. It is customary to omit the decimal point and say that Harold has an intelligence quotient<sup>2</sup> (IQ) of 110. Similarly Frank's IQ would be  $\frac{10}{10}$  or simply 100, and George's intelligence quotient would be  $\frac{9}{10}$  or 90. We may think of

<sup>1</sup> A term used to mean age in the ordinary sense as distinguished from mental age.

<sup>2</sup> In finding an intelligence quotient, it is customary to consider all persons of more than 16 years as being just 16, for reasons that will be explained.

this value of 90 as meaning 90 per cent of normal in mental age.

It has been found that intelligence quotients thus obtained by means of the Stanford Revision of the Binet-Simon Test are distributed approximately according to the law of normal distribution for each age and in such a manner that the variability of intelligence quotients is approximately the same for the various ages. In other words, about 25 per cent of children of each age have intelligence quotients of 108 or more, and about 75 per cent of children of each age have intelligence quotients of 92 or more, as we saw in Chapter IX. The significance of the various intelligence quotients has been stated by Dr. Terman on page 79 of *The Measurement of Intelligence* <sup>1</sup> in Table 28.

TABLE 28  
DR. TERMAN'S CLASSIFICATION BY IQ

IQ	CLASSIFICATION
Above 140 . . . . .	"Near" genius or genius.
120-140 . . . . .	Very superior intelligence.
110-120 . . . . .	Superior intelligence.
90-110 . . . . .	Normal, or average, intelligence.
80- 90 . . . . .	Dullness, rarely classifiable as feeble-mindedness.
70- 80 . . . . .	Border-line deficiency, sometimes classifiable as dullness, often as feeble-mindedness.
Below 70 . . . . .	Definite feeble-mindedness.

This interpretation has now come to be widely accepted as standard, and the IQ, as the standard measure of brightness, is now familiar to practically all who make use of mental-ability tests. Brightness may be measured in terms of a percentile rank, and numerous other measures of brightness

<sup>1</sup> Houghton Mifflin Company.

have been devised for use with various other tests and for various purposes. These are called by different names, such as coefficient of intelligence, coefficient of brightness, index of brightness, etc. These are discussed below.

**The invalidity of the old IQ.** By the term "old IQ" is meant the IQ as usually derived from a mental age either by the Binet-Simon Test or a group test, found by dividing that mental age by the chronological age (with 16 years as a maximum) as distinguished from what we might call the new IQ, by which it is proposed to supplant the old. As suggested in the paragraph heading, the IQ as customarily found at present is subject to certain inconsistencies for the reasons that follow.

First let us consider the case of the IQ as found from the Binet-Simon Test. The mental-age scale of the Binet-Simon Test, as we have already seen, really represents scores expressed in terms of years and months. These are normal for the corresponding ages in the lower part of the scale but are fictitious in the upper part. Let us assume, however, that the units of the scale are equal throughout — that the difference in mental ability between the mental ages of 10 and 11 years is the same as the difference between mental ages of 11 and 12, etc., all the way up to 19 years, 6 months, the top of the scale as shown in Figure 45.

Let us assume that the normal curve of growth in mental ability in terms of mental ages by the Binet-Simon Test is as shown in Figure 45. In this we assume the mental ages to be true mental ages up to the age of 14 years and that normal mental maturity is represented by a Binet mental age of 16 years. This is the assumption underlying the present practice of using 16 years as the maximum chronological age when calculating an IQ.

We could not assume the mental ages to be true mental ages up to 16 years without drawing the line straight up to

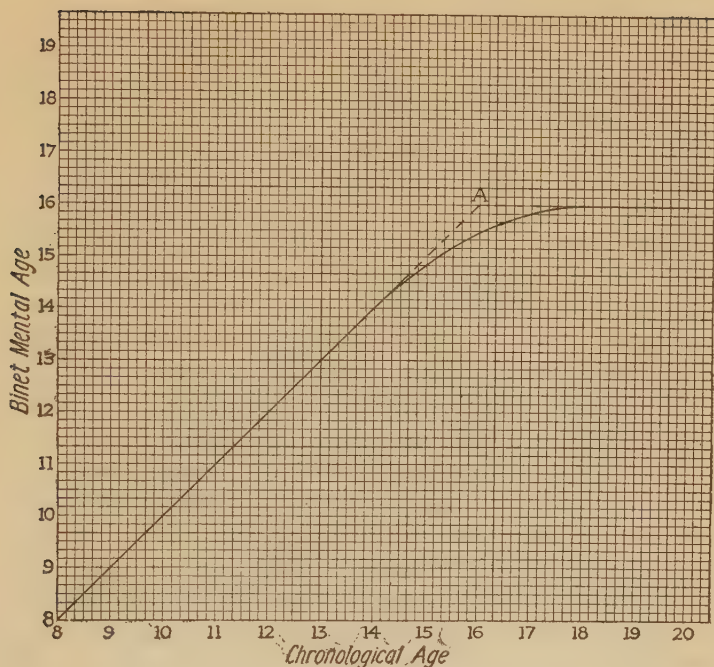


FIG. 45. The hypothetical normal curve of growth in mental ability in terms of mental ages by the Binet-Simon Test.

the level of a Binet mental age of 16 years (point *A*) and then making a sudden jog off at a level, as shown by the broken lines. All the evidence is against any such assumption as this.

Now the figure shows plainly that the normal score (Binet mental age) for the age of 16 must be somewhat below 16 years. According to our hypothetical curve it would be about 15 years, 5 months. A 16-year pupil with a Binet mental age of 15 years, 5 months would be exactly normal and should have an IQ of 100. By the present method of figuring, however, his IQ is  $\frac{15 \text{ years, } 5 \text{ months}}{16 \text{ years}} = 97$ . You

can see, of course, that it would not be 100 under these conditions.

We might assume, on the other hand, that the curve really was straight up to 16 years (point *A*), in which case, of course, the IQ of a normal 16-year pupil would come out just 100. But in that case the curve must extend beyond a Binet mental age of 16 years because we know that it cannot form an angle at that point. That would mean that the normal Binet "score" for adults is some mental age above 16 — let us say 16 years, 6 months. In that case the IQ of a normal adult found in the customary manner would be 
$$\frac{16 \text{ years, 6 months}}{16 \text{ years}} = 103.$$
 You can see here, too, of course

that it could not be 100 under these conditions.

We see from these facts that the IQ of the normal individual, calculated as is customary at present, must either fall somewhat below 100 at some point in the scale or it must rise above 100 at some point, or both.

**The remedy.** The remedy for this difficulty is very simple, as far as the Binet-Simon Test is concerned. It consists merely in using for the denominator of the fraction when figuring the IQ, not the chronological age of the individual but the Binet mental age that is the norm for his age. The IQ of the normal 16-year pupil is then 
$$\frac{15 \text{ years, 6 months}}{15 \text{ years, 6 months}} = 100,$$

as it should be, and the same for every other age. The IQ of a truly normal individual will always be found to be exactly 100, provided the norms are correct.

Unfortunately, however, there are at present no age norms in the Binet-Simon Test. No normal curve of growth in mental ability has been determined. We do not know what Binet mental age is normal for 16-year-olds. Guesses are being made that it is somewhere between Binet mental ages of 14 and 16 years, but the statistical determination yet remains to be made.

**A popular misconception.** It will be well at this point to call attention to an error frequently made in confusing the age at which mental maturity is reached with the Binet mental age representing normal mental maturity. According to our hypothetical curve of growth, normal mental maturity is represented by a Binet mental age of 16 years, but it is not reached by the normal individual until the age of 18 years. Obviously it is conceivable that normal mental maturity might be represented by a Binet mental age of 14 years (some claim this is the case), and that it is not reached by the normal individual until the age of 19 years.

Do not make the mistake, therefore, as many have, of thinking that when a person says he believes "the mental level of the average adult is a Binet mental age of 15 years" he means or implies that individuals reach mental maturity at the age of 15 years.

A still more serious error would be to assume that if an individual reaches maturity at the age of 18 years, the normal mental ability of adults is represented by a Binet mental age of 18 years. Just such errors as that are made by some who contribute popular articles on the subject of mental testing to lay magazines.

**The Coefficient of Brightness.** In order to obtain a measure of brightness that would not be subject to the inaccuracies inherent in the IQ as customarily obtained at present and which, nevertheless, would be comparable with it, the writer devised a measure of brightness for use with the Otis Group Intelligence Scale, Advanced Examination, which was called the Coefficient of Brightness (CB). It happened that the norms as at first derived for this test were such that the score plus 55 points equaled the number of months in the corresponding Binet mental age. Instead of dividing a pupil's Binet mental age, so found, by his chronological age to find an IQ, it was divided by the Binet mental age that was the

norm for the pupil's age. This was found by adding 55 points to the score which was the norm for his age. The result was a measure of brightness comparable to the IQ in general and not subject to the errors described above that enter into the IQ as now generally obtained.

This method is now used in the case of the Illinois Examination but has been displaced in the case of the Otis Group Intelligence Scale by a measure of brightness called the Index of Brightness (IB), which, while not directly comparable with the IQ, is more constant and therefore more valid than either the IQ or Coefficient of Brightness for use with a group test, as is shown in the following pages.

**Intelligence quotients from group tests.** Thus far we have been considering the validity of the IQ as obtained by the Binet-Simon Test. It is possible, of course, to obtain an IQ of any individual who has been tested by a group test of mental ability by finding the Binet mental-age equivalent of his score in the group test and dividing this by his chronological age as usual, or by dividing the Binet mental age equivalent of his score by the Binet mental age that is normal for his chronological age—this method being the one recommended for finding the “new IQ.” It has been found, however, that intelligence quotients obtained from group tests by either of these methods vary more widely for young children than for older children. This is apparently due to the fact that the scores of young children vary more widely in proportion to the ages of these children than do the scores of older children.

On account of these facts the intelligence quotient as a measure of brightness loses much of its significance when used with group tests, since an IQ of 120, for example, denotes a much higher degree of brightness in the case of an older child than it does in the case of a younger child, whereas, in order to be a valid measure of brightness, an intelligence

quotient at one age should represent the same degree of brightness that it does at other ages.

Not only do the scores of younger pupils in group tests vary more in proportion to the age of these pupils than do the scores of older pupils, but they vary to practically the same extent as do the scores of older pupils. Indeed, distributions of scores in the National Intelligence Test <sup>1</sup> and various other mental-ability tests have been found to be approximately equally variable for both old and young pupils.

These facts suggest the need of a method of expressing brightness which is different from that of finding the intelligence quotient of a pupil by dividing his mental age by his chronological age. The new method should be such that a pupil having a percentile rank of 75 at the age of 8 would have the same measure of brightness as a pupil having a percentile rank of 75 at the age of 16.

Now it has been found that a pupil having a percentile rank of 75 in mental ability among pupils of his age will make a score in the Otis Group Intelligence Scale, Advanced Examination, about 20 points above the norm for his age, no matter at what age. And similarly a pupil having a percentile rank of 25 will make a score of about 20 points below the norm for his age, no matter at what age. These facts suggest that, rather than find a measure of brightness by division, it would be better and simpler to find it by subtraction. That is, it would seem that the number of points by which a pupil exceeds or falls below the norm for his age is a more constant measure of brightness than an IQ obtained in the usual manner.

**The Index of Brightness.** It is not convenient, of course, to express brightness by negative numbers, but we can easily avoid this by letting 100 represent exact normality the same as in the case of the IQ, and adding to or subtracting from

<sup>1</sup> See Supplement No. 3 to the Manual of Directions.

100 the number of points by which a pupil's score exceeds or falls below the norm for his age.

Provision has been made in the case of the Otis Group Intelligence Scale for expressing brightness in just those terms, and the measure is called the Index of Brightness.

**A new method of finding an intelligence quotient.** In the case of the Higher and Intermediate Examinations of the Otis Self-Administering Tests of Mental Ability the same fairly constant variability of scores from age to age was found. This called for a method of expressing brightness similar to the Index of Brightness.

Fortunately it happened that the variability of scores of both tests at each age was such that the median deviation of the distributions was approximately 8 points — just the value of the median deviation of IQ's by the Binet-Simon Test! This meant that the Index of Brightness by either the Higher or Intermediate Examination was directly comparable with the IQ by the Binet-Simon Test. That is, an Index of Brightness of 108 denoted a percentile rank of 75 exactly as a Binet IQ of 108 did, and the same for other values.

This being the case, it seemed best simply to call this new measure of brightness an "IQ" (although it is not a quotient, of course), because the term IQ is so widely used and because this measure of brightness is seemingly more closely related to Binet IQ's than actual quotients found by dividing Binet mental-age equivalents of scores in group tests by chronological age.

**Percentile rank.** In our discussion of the validity of measures of brightness we have referred frequently to percentile rank of a pupil in mental ability among pupils of his own age as the best criterion of brightness. And indeed it is. Its only drawback is the fact (which we have considered) that units of percentile rank are not of equal value throughout the

scale. It may be also that we have learned to prefer 100 rather than 50 as a measure of normality. But it is perhaps the most easily understood and interpreted of all the measures and no doubt will be used more in the future than in the past.

**The Interpretation Chart.** In order to facilitate the calculation of IQ's from pupils' scores in the Intermediate and Higher Examinations, an interpretation chart has been drawn for each of these tests. The Interpretation Chart for the Intermediate Examination is shown in Figure 46. Note the age scale at the foot of the chart, the score scale at the right, and the heavy curve through the middle of the chart. This heavy line is the curve of growth of the exactly normal child in mental ability. The curve shows that the norm for the age of 8 years is 7 points, the norm for the age of 9 years is 15 points, and so on up to 17 years, for which the norm is 58 points.

**Finding an intelligence quotient by means of the chart.** Let us suppose a pupil is 12 years, 10 months old. The heavy curved line cuts the vertical line above 12:10 on the age scale at a point on the horizontal line representing a score of 43. This means that the norm for the age of 12 years, 10 months is 43. No account need be taken of this, however, in finding an IQ. Thus let us suppose that the pupil's score in the Intermediate Examination was not 43, but 49 points. We should then find the point on the vertical line representing the age of 12 years, 10 months, and on the horizontal line the point representing the score of 49. It will be seen that this point falls on one of the curves, and if we follow this curve to the right side of the chart we find the number 106. This means that the pupil's IQ is 106.

In one operation the chart determines the norm for the pupil's age, finds the number of points by which the pupil's score exceeds the norm for his age, and adds this amount to 100. It is not necessary to find a mental age, or to convert an

OTIS SELF-ADMINISTERING TESTS OF MENTAL ABILITY

INTERPRETATION CHART. For Intermediate Examination

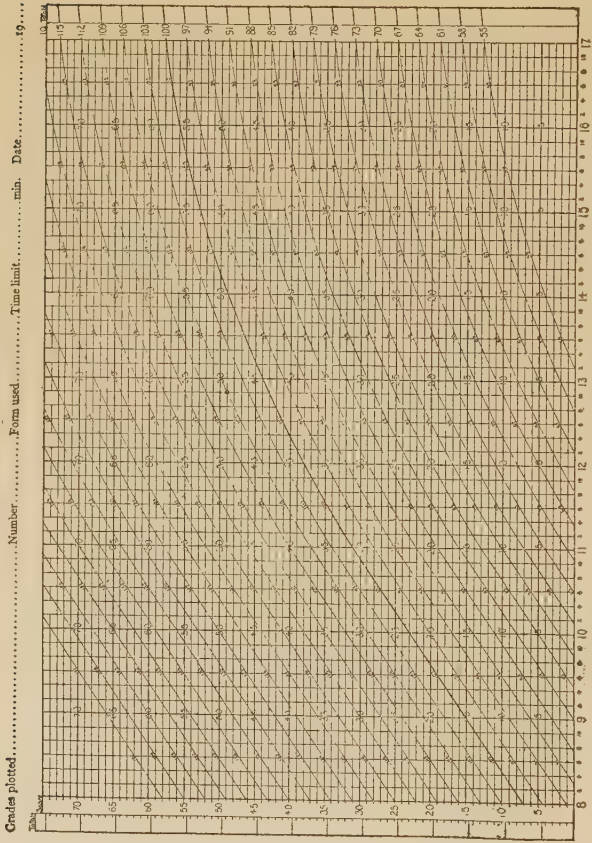


FIG. 46. Interpretation Chart for Intermediate Examination. (Actual size, 7 by 10 inches.)

age expressed in years and months to terms of months only, or to carry out any division.

Provision is made on the chart so that the point representing the score and the age of each pupil may be entered as a dot on the chart and the distribution of IQ's of a group found by simply adding the dots that fall between each pair of adjacent curves.

#### QUESTION

Which of the various methods of measuring brightness described in this chapter do you prefer?

## CHAPTER FOURTEEN

### NORMS

THUS far we have dealt with norms only as age norms — scores in a test that are normal for each age. There are also grade norms — scores in a test that represent the median ability of the pupils of each grade. We can have mental-age norms also if we wish. The score in a test that is the norm for the mental age of 10 years would be the median score of all pupils whose mental age is 10 years. Mental-age norms, then, are the scores in achievement tests that are normal for the various mental ages.

**Need for representative sampling.** Pupils differ so much in ability from group to group that in order to base the norms on a really representative sampling of pupils it is necessary to have scores of at least a thousand pupils of each age, and these should be from a wide range of territory and should include both city and rural schools.

**Age norms.** The great difficulty in obtaining age norms is that of getting a really unselected group at each age. If we wish to find the median score of 8-year children in a test, we should include the 8-year pupils in the first and second grades as well as those in the third grade and above. We should also include, of course, all 8-year children who have not yet entered school. This is usually impossible, so that we can only estimate how many there are of such pupils and what their scores would be.

Similarly, if we wish to find the median score of 17-year pupils in a test we should include not only those 17-year pupils who are in school but those who have dropped out of school and those who have graduated. This also is ordinarily impossible, so that here again we can only guess as to how many there are of these and what their scores would be.<sup>1</sup>

<sup>1</sup> The method of estimate used in the case of these lower and upper ages is well illustrated in the description of the derivation of norms for the Otis

**Grade norms.** There are so many factors that cause the median scores of this and that grade or school or city to vary from similar medians in other parts of the country that grade norms, no matter how carefully obtained, are not of great value. Among the many conditions causing median scores of grades and schools to vary are the following:

(a) When the pupils of a grade or a school are being tested for the first time, some of the pupils may fail to understand just what is wanted — how to do the underlining, etc. The many distractions prevent them from doing as well as they could do if they were quite familiar with tests and knew just how to go about them. Degree of familiarity with standard tests, then, affects the median score of a class.

(b) Examiners differ much in the way in which they administer tests. A principal or a teacher who is revered by the pupils may stimulate them to sustained and earnest thought upon a test, while another may make remarks which cause them to be over-anxious.

(c) The accuracy of scoring varies much from school to school. The writer once checked over a bunch of test papers that had been scored by teachers and in more than half the papers found from one to twenty-five errors in the scoring of seventy-five questions! The carelessness is nearly always on the side of overlooking errors of the pupils. The median score of a class in such a case would be materially raised above what it should be.

(d) Again, it has been found that the average age of pupils of a certain grade varies very markedly from city to city and even from school to school within a city. A variation of one whole year in median age of a grade is not uncommon within one school system. Obviously a fifth-grade norm

Group Intelligence Scale beginning on page 54 of the Manual of Directions, 1921 Revision.

based on a median age of eleven years cannot apply to a fifth grade with an average age of ten years.

(e) The median scores of the various grades of a school or a school system in a mental-ability test will be found to be affected to a large extent by the social status of the community. It appears to be possible for the median score of all fifth grades, for example, in one city to be as high as the median of the sixth grades of another city.

(f) Median scores in achievement are affected not only by differences in social status of the community but in quality of instruction also, and it is well known that quality of instruction varies enormously from school to school and system to system.

Considering the many causes of variation in median scores of different groups of pupils, you will see that, after all, norms are not applicable to all such groups, and you should not feel worried if the scores of your pupils are not exactly in accord with the norms given for a certain test.

It is assumed that what a teacher wishes to know about her pupils is first of all whether they are as bright as the pupils in the other grades or in the same grade in other schools of the city and if so whether they have learned as much for their grade. It is assumed that a teacher wishes to know next whether the pupils of her class vary markedly in mental ability and achievement; who are the bright and who are the dull pupils in her class, and which pupils are not accomplishing as much as they should in school work and why.

It is assumed that what the principal wishes to know is whether the pupils of his school are as bright as those of other schools, and if so whether they are accomplishing as much, whether the pupils are well graded so that they can be most efficiently taught and if not which pupils are misplaced, etc.

All these facts can be determined without grade norms. Age norms tell who are bright and who are dull, median scores of classes show how pupils compare by groups, percentile curves show the variability of scores and the overlapping of grades, and comparison of scores in mental ability and achievement with age norms shows the degree to which a child is accomplishing what he should. The reason for the failure of any particular child is to be discovered by diagnostic tests.

One of the purposes of this book is to help teachers and principals to go farther than merely to test the pupils, compare the median scores with grade norms, determine that the grade or school is above or below "standard," and file the papers.

Of course there is no harm in comparing a grade median with a grade norm which constitutes the median score of all the pupils of a given grade which the test makers can bring together. Such comparison, however, should be considered merely incidental and of much less importance than the comparison with median scores of other classes in the same school or city.

**How to obtain grade norms from age norms.** If it is desired to compare grade medians with grade norms for a test for which no grade norms are furnished or in case the norms furnished do not seem to apply, grade norms for mid-year may be derived from the age norms on the basis of the median grade of pupils of the various ages as obtained from the investigation of the age and grade of a million children.

Table 29 shows the percentage distribution of a million children by age in each of Grades 1 to 12, based on 80 cities. This table was compiled by H. R. Bonner of the Bureau of Education at Washington, in 1918.

The median age of pupils of each grade as determined from Table 29 is shown in Table 30.

TABLE 29  
PERCENTAGE DISTRIBUTION OF A MILLION PUPILS BY AGE, IN EACH GRADE, BASED ON 80 CITIES, 1917-18

AGES LAST BIRTHDAY — YEARS																		
GRADES																		TOTAL
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21+	
1	10.30	54.91	24.59	6.84	2.07	.73	.28	.13	.08	.04	.02	.01						100
2	.15	9.25	44.87	28.35	10.65	4.16	1.45	.67	.28	.11	.05	.01						100
3		.27	9.71	38.78	28.36	13.13	5.45	2.52	1.14	.47	.14	.02	.01					100
4			.38	9.29	35.12	27.49	14.58	7.55	3.63	1.46	.42	.06	.01	.01				100
5			.01	.56	10.26	32.27	26.58	15.98	9.34	3.67	1.14	.17	.02					100
6				.04	1.00	11.01	31.87	27.55	17.55	7.98	2.60	.34	.05	.01				100
7					.03	1.24	12.03	33.58	30.83	15.69	5.32	1.08	.16	.03	.01			100
8						.07	1.53	14.14	37.20	30.09	13.12	3.21	.51	.09	.02	.01	.01	100
9							.10	2.61	17.45	37.00	26.88	11.31	3.48	.82	.21	.07	.07	100
10							.01	.19	3.15	18.61	36.34	26.43	10.99	3.18	.71	.22	.17	100
11								.02	.38	3.29	18.83	34.89	28.00	10.59	3.07	.66	.27	100
12									.02	.28	3.56	18.38	36.70	27.16	10.11	2.89	.90	100
Avg.	1.68	10.04	10.70	10.49	10.57	10.32	9.82	9.58	9.28	7.19	4.67	2.76	1.73	.81	.26	.07	.03	100

TABLE 30

SHOWING THE MEDIAN AGE OF PUPILS OF EACH GRADE BASED ON  
A MILLION CHILDREN FROM 80 CITIES

GRADE	MED. AGE	GRADE	MED. AGE
12	17:9	6	12:2
11	16:9	5	11:2
10	15:10	4	10:1
9	14:11	3	9:0
8	14:0	2	7:10
7	13:1	1	6:8

The median age of sixth-graders is 12 years, 2 months. If we assume that the pupil of median age in any grade is just normal, then the norm in any test for the age of 12 years, 2 months is the norm for Grade 6 in the middle of the year.

This reasoning would not hold true for the lowest and highest grades, because in the one case dull pupils have not yet entered and in the other dull pupils have dropped out. Between Grades 3 and 10, however, the reasoning is probably approximately correct.

In these cases, then, we may find grade norms from the age norms as follows. To find the norm for Grade 6 at mid-year look up the norm for 12 years, 2 months. And the same for the other grades, following Table 30.

It may be contended that theoretically the "median grade" for each age should have been used as the basis for finding the relation between age and grade for this purpose. It happens, however, that when this method is used the relation between age and grade is almost exactly the same as that given in Table 30 and for that reason we need not be concerned about these theoretical considerations.

**Grade status.** It is becoming the custom to express the scores of pupils in terms of grade norms more precisely than by whole-year steps. Thus, we might let 6.1 denote the normal achievement of pupils in the sixth grade at the end

of the first month of the school year, let 6.2 denote the normal achievement at the end of the second month, etc. On the basis of a school year of ten months, which is taken as standard,<sup>1</sup> 6.5 would then denote the normal achievement of sixth-grade pupils at the middle of the year. We may say, then, that the score in a test that is normal for age 12 years, 2 months, represents a grade status (G.S.) of 6.5; that the score that is normal for age 13 years, 1 month, represents a G.S. of 7.5; etc., according to Table 30.

The term "*G* score" is sometimes used to denote what we have called grade status. Also the term "*B* score" has been used in this way, but this is unfortunate for the reason that this term was originally used by Professor McCall to refer to measures of brightness, and its use to refer to grade status is sure to lead to confusion. It seems obvious that the letter *B* should be reserved for measures of brightness and that the letter *G* is more suited for measures of grade status.

An alternative method of expressing grade status is to let 3.1, for example, represent the normal achievement of the middle of the first month of the third grade, instead of the end of the first month. This method might be desirable when a 9-month year is used as standard in order that the grade status corresponding to the middle of the year will be 3.5, 4.5, etc. However, since a mental age of 3 years, 1 month, for example, means the mental age of a normal child at the *end* of the first month after the birthday, so, to be consistent, it seems preferable to let 3.1 stand for grade status at the *end* of the first month of the third grade; 3.0 will then represent grade status at the beginning of the first month of the third grade or the end of the 10th month of the second grade.

A scale of grade status, therefore, may be made up from

<sup>1</sup> No serious consequences result from the use of this basis even in cases where the term is somewhat less than ten months.

age norms by means of Table 31, in which intervening values have been supplied by interpolation.<sup>1</sup>

TABLE 31

FOR FINDING THE GRADE STATUS OF A PUPIL FROM AGE NORMS. A GRADE STATUS OF 3.0 IS REPRESENTED BY THE NORM FOR THE AGE OF 8 YEARS, 6 MONTHS, ETC.

GRADE STATUS	AGE FOR WHICH IT IS THE NORM	GRADE STATUS	AGE FOR WHICH IT IS THE NORM	GRADE STATUS	AGE FOR WHICH IT IS THE NORM	GRADE STATUS	AGE FOR WHICH IT IS THE NORM
3.0	8:6	5.0	10:7	7.0	12:8	9.0	14:5
3.1	8:7	5.1	10:8	7.1	12:9	9.1	14:6
3.2	8:8	5.2	10:10	7.2	12:10	9.2	14:7
3.3	8:10	5.3	10:11	7.3	12:11	9.3	14:8
3.4	8:11	5.4	11:0	7.4	13:0	9.4	14:10
3.5	9:0	5.5	11:2	7.7	13:1	9.5	14:11
3.6	9:2	5.6	11:3	7.6	13:2	9.6	15:0
3.7	9:3	5.7	11:4	7.7	13:3	9.7	15:1
3.8	9:4	5.8	11:5	7.8	13:4	9.8	15:2
3.9	9:6	5.9	11:6	7.9	13:5	9.9	15:3
4.0	9:7	6.0	11:8	8.0	13:6	10.0	15:4
4.1	9:8	6.1	11:9	8.1	13:7	10.1	15:5
4.2	9:9	6.2	11:10	8.2	13:8	10.2	15:6
4.3	9:10	6.3	11:11	8.3	13:9	10.3	15:7
4.4	10:0	6.4	12:1	8.4	13:11	10.4	15:9
4.5	10:1	6.5	12:2	8.5	14:0	10.5	15:10
4.6	10:2	6.6	12:3	8.6	14:1	10.6	15:11
4.7	10:3	6.7	12:4	8.7	14:2	10.7	16:0
4.8	10:5	6.8	12:6	8.8	14:3	10.8	16:2
4.9	10:6	6.9	12:7	8.9	14:4	10.9	16:3

Finding grade status from age norms. By way of application let us suppose that it is desired to make up a table of grade status corresponding to scores in the Stanford Achievement Test, Advanced Examination. With the help

<sup>1</sup> See page 169 for the meaning of interpolation.

of Table 31 this may be done very easily. Thus, to find the score representing a grade status of 3.0 we have but to look up the norm for the age of 8 years, 6 months. This is a score of 13. The score representing a G.S. of 3.1 is the norm for the age of 8 years, 7 months, etc., according to Table 31. This is a score of 14. In this way a complete table of grade status could be made up.<sup>1</sup>

**Finding age norms from grade norms.** It sometimes happens that a test is furnished with grade norms but without age norms, since age norms are of little value in the case of subject-matter tests except for finding accomplishment ratios. (See page 174 for an explanation of accomplishment ratios.) In such a case age norms may be obtained from the grade norms by means of the same table we used for finding grade status from age norms (Table 31). The manual containing the Hudelson English Composition Scale<sup>2</sup> gives grade norms in quality of composition for Grades 4 to 12. No age norms are given, since until recently the interest in norms in achievement tests has been to show how a pupil compared in ability with others in his grade.

Let us assume, however, that it is desired to express the scores of pupils in quality of composition in terms of age norms. It should be understood that this is not a procedure that is altogether safe for general application, for if we are considering geography, for example, you will see, of course, that the gain of pupils in knowledge of geography is far greater in the grade in which they are studying the subject than in grades above or below it, whereas the method herein described rather assumes a fairly uniform improvement from grade to grade. This would be more or less true of ability in composition, so we may more reasonably apply the method.

<sup>1</sup> The grade-status values furnished by the authors of this test, based upon only those cases from which the norms were derived, differ slightly from the values obtained in this way.

<sup>2</sup> Published by World Book Company, Yonkers-on-Hudson, New York.

The grade norms for January in the Hudelson English Composition Scale are as follows :

Grade . . . . .	4	5	6	7	8	9	10	11	12
Norm (Quality score)	3.0	3.6	4.2	4.7	5.3	5.5	5.9	6.3	6.7

From Table 31 we can find the ages for which these grade norms (quality scores) are normal. Assuming January to be the 5th month, we look for the ages opposite 4.5, 5.5, etc., in the table. We find the ages for which these scores are the norms to be as follows (see Table 30 for upper ages) :

Quality scores . .	3.0	3.6	4.2	4.7	5.3	5.5	5.9	6.3	6.7
Ages for which normal . . .	10:1	11:2	12:2	13:1	14:0	14:11	15:10	16:9	17:9

How, now, shall we find what quality score is the norm for the age of 10:2, 10:3, etc.? This is done by the method of interpolation.

**Interpolation.** This term applies to the finding of intervening values in a scale, as, for example, the normal quality scores for the ages between 9:0, 10:1, 11:2, etc., in the Hudelson English Composition Scale.

The method is to plot points on a chart and to draw a line showing the relation between age and score. Thus, in Figure 47 the horizontal scale represents ages and the vertical scale represents quality scores. The position of the lowest point (circle) shows that the quality score of 3.0 is the norm for the age of 10 years, 1 month. The next point shows that the score of 3.6 is the norm for the age of 11 years, 2 months, etc.

A smooth curve has been drawn through these points with one exception, and it would appear from the position of this point that the norm for Grade 8 as given is slightly too high.

For purposes of interpolation we assume that as the normal pupil grows older his quality score in composition tends to in-

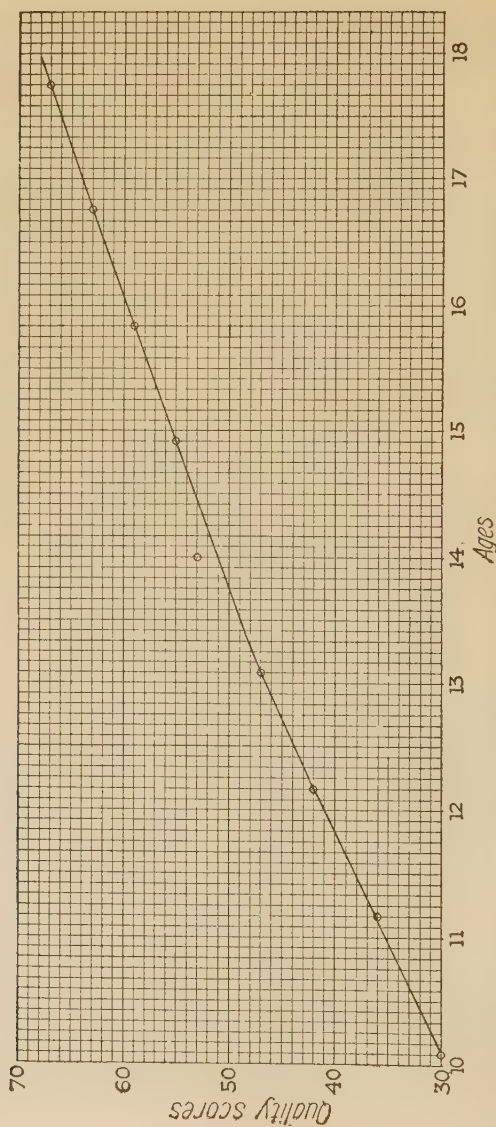


FIG. 47. Showing the method of deriving age norms, in the Hudelson English Composition Scale, from grade norms by the method of interpolation.

crease in accordance with the line. This line, in other words, is the "curve of growth" of a normal child in composition.

We have now merely to read from the line the intervening values of scores — those corresponding to the ages of 10:2, 10:3, etc. These are as follows:

Age . . . . .	10:2	10:3	10:4	10:5	10:6	10:7	10:8 etc.
Norm (Quality Score)	3.0	3.1	3.1	3.2	3.2	3.3	3.3 etc.

In this way we can build up a table of age norms.

**Subject ages.** It is becoming the custom to express the score of a pupil in an achievement test in terms of the age for which the score is the norm. Thus the age for which the quality score of 3.1 is the norm is 10 years, 3 months; so we may say, if we wish, that a pupil making a score of 3.1 has a *composition age* of 10 years, 3 months. A pupil making a score in arithmetic that is normal for the age of 10 years, 3 months we may say has an *arithmetic age* of 10 years, 3 months, etc. These various ages are called *subject ages*.

**Fictitious subject ages.** Just as some persons make scores in mental ability much above the norm for adults, so do they often make scores in achievement tests in the school subjects much above the norm for adults. For that reason and because of the habit we have formed of thinking of mental ability in terms of mental age, many wish to think of achievement in school subjects in terms of subject age even though the achievement is beyond that normal for adults. For this reason scales of subject ages are often extended beyond the norm for adults in such a way that the subject ages become fictitious in just the same way that mental ages do, as previously explained (see page 146).

**Educational quotient.** Just as the mental age divided by the chronological age is called the *intelligence quotient*, so the reading age of a pupil divided by his chronological

age is called his *reading quotient*. His arithmetic age divided by his chronological age is called his *arithmetic quotient*, etc. Each of these quotients is called a *subject quotient*. The average of a pupil's subject quotients is called his *educational quotient*. Or his educational quotient may be found by dividing his achievement age as a whole by his chronological age. Thus a 10-year pupil whose reading age is 11 years has a reading quotient of  $\frac{11}{10}$ , or 110 (that is, 110 per cent of normal). If his arithmetic age is 9 years, he has an arithmetic quotient of  $\frac{9}{10}$ , or 90 (that is, 90 per cent of normal). If this pupil's achievement age (average subject age or age equivalent in a general achievement test) is 10 years, his educational quotient is  $\frac{10}{10}$ , or 100.

**The accomplishment ratio.** Dr. Raymond Franzen<sup>1</sup> has popularized the ideal of making a comparison between a pupil's subject age as in composition or reading, and his *mental age*, to determine the degree to which he was achieving what we have a right to expect from a pupil of his mental age. If a pupil has a mental age of 10 years, we have a right to expect him to have a subject age of 10 years in each of the school subjects, no matter how old he may be. If he has a reading age of 11 years, he is really accelerated in reading (even though he may be 11 years old) and we may divide his reading age by his mental age and say he has a *reading ratio* of  $\frac{11}{10}$ , or 110 (110 per cent of the reading ability normal for his mental age). If a pupil with a mental age of 10 years has a spelling age of only 9 years, he is really retarded in spelling (even though he is only 9 years old), and we may say that he has a *spelling ratio* of  $\frac{9}{10}$ , or 90 (90 per cent of the spelling ability normal for his mental age).

The average of a pupil's *subject ratios* (S.R.), or the ratio of his *achievement age* as a whole to his mental age (which

<sup>1</sup> Raymond Franzen, "The Accomplishment Quotient," *Teachers College Record*, November, 1920.

amounts to the same thing), is called his *accomplishment ratio* (A.R.).<sup>1</sup>

**Summary.**

$\frac{\text{Arith. Age}}{\text{Chron. Age}} = \text{Arith. Quot.}$	$\frac{\text{Arith. Age}}{\text{Mental Age}} = \text{Arith. Ratio}$
$\frac{\text{Read. Age}}{\text{Chron. Age}} = \text{Read. Quot.}$	$\frac{\text{Read. Age}}{\text{Mental Age}} = \text{Read. Ratio}$
$\frac{\text{Subject Age}}{\text{Chron. Age}} = \text{Subject Quot.}$	$\frac{\text{Subject Age}}{\text{Mental Age}} = \text{Subject Ratio}$
$\frac{\text{Achieve. Age}}{\text{Chron. Age}} = \text{Ed. Quot.}$	$\frac{\text{Achieve. Age}}{\text{Mental Age}} = \text{Accomp. Ratio}$
$\frac{\text{Mental Age}}{\text{Chron. Age}} = \text{Intel. Quot.}$	$\frac{\text{Ed. Quot.}}{\text{Intel. Quot.}} = \text{Accomp. Ratio}$

The accomplishment ratio can be found by dividing the achievement age by the mental age when both of these have been found on approximately the same date; otherwise it is necessary to find the accomplishment ratio by dividing the educational quotient by the intelligence quotient. The lower fraction is the same as the upper, with the numerator and denominator each divided by chronological age, and when the latter is the same in both cases this does not change the value of the fraction.

We have already seen in Chapter XIII how a measure of brightness which has been called the "IQ" can be found by subtraction instead of division in the case of the Otis Self-Administering Tests of Mental Ability. In the case of the Otis Classification Test, which is a combination of the Intermediate Examination and an Achievement Test, it is possible to find an "Educational Quotient" in the same way, and by a comparison of scores in the Mental Ability Test and the

<sup>1</sup> This was first called the accomplishment quotient.

Achievement Test an "Accomplishment Ratio" may be found by the subtraction method.

In a recent communication to the writer Dr. Franzen states that he believes the method of finding measures of the relation of achievement to mental ability by *division* (i.e., dividing educational age or subject age by mental age) involves "important statistical limitations" and that he has adopted a method of subtraction for finding this relation. Obviously subtraction is easier than division, and there seems to be no special reason why subtraction should not be used for this purpose.

# CHAPTER FIFTEEN

## THE MEANING OF CORRELATION

**The comparison of relative standing.** Let us suppose that there are 64 pupils in a certain group for whom a teacher has given the ratings A, B, C, D, and E, for the month, according to her estimate of their daily work in arithmetic. Let us suppose that these pupils have been given a test of ten problems in arithmetic covering the work of the month and we wish to compare their relative standing in the test and in their daily work according to the teacher's rating.

Suppose that the ratings and scores of the 64 pupils are as shown in Table 32.

TABLE 32

SHOWING THE HYPOTHETICAL TEACHER'S RATINGS OF 64 PUPILS IN ARITHMETIC AND THEIR HYPOTHETICAL SCORES IN AN ARITHMETIC TEST

PUPIL	RAT- ING	SCORE	PUPIL	RAT- ING	SCORE	PUPIL	RAT- ING	SCORE	PUPIL	RAT- ING	SCORE
1	A	10	17	C	8	33	C	9	49	E	7
2	D	9	18	D	9	34	C	7	50	C	7
3	B	8	19	B	7	35	B	10	51	D	6
4	C	9	20	C	9	36	D	7	52	B	8
5	C	8	21	A	9	37	C	8	53	C	8
6	B	9	22	D	8	38	C	8	54	D	7
7	C	6	23	B	10	39	E	7	55	B	9
8	D	8	24	C	10	40	B	8	56	C	7
9	A	8	25	D	8	41	B	8	57	B	9
10	C	8	26	C	8	42	C	9	58	D	7
11	E	6	27	B	8	43	D	6	59	C	8
12	B	9	28	C	8	44	D	8	60	C	7
13	C	9	29	A	9	45	B	9	61	C	9
14	E	8	30	C	7	46	C	7	62	D	7
15	C	8	31	D	8	47	D	7	63	D	8
16	B	8	32	B	9	48	D	7	64	B	7

**The meaning of correlation.** When we say that we wish to compare the relative standing of pupils in daily work and in the test we mean, of course, that we wish to see whether those pupils who did well in their daily work also did well in the test and whether those who did poorly in one did poorly in the other, and to what extent this was true. In other words, was there a marked tendency for the pupils who did well in daily work to do well in the test also and to do as well in the test as in daily work, or was there only a slight tendency and how slight or how marked? If we should find that there was a tendency for those who got good ratings in daily work to get good scores in the test also, and vice versa, we should say that the test scores *correlated* with the teacher's ratings. If there were a very marked tendency, we should say there was a *high correlation* between scores and marks; and if the tendency were only a slight tendency, we should say there was a *low correlation*.

**The natural method.** The most natural way to find the degree to which pupils who get high ratings tend to get high scores also is to separate the pupils into groups according to ratings and see what scores those made who got A, what scores those made who got B, etc.

If we have each pupil's rating and score on a card, we can sort the cards into piles according to rating, putting all the A's in one pile, the B's in another, etc. This is the easiest and safest way, for we can then run through each pile and make sure that we did not get any B's in the A pile, etc.

But if the ratings and scores are not on cards, we can distribute them in a table, putting the A's in one column, the B's in another, etc., and distributing the scores of each column by a scale, as shown in Table 33. (There is an artificial symmetry about this table that has been introduced for the sake of convenience in subsequent calculation.) Any table of this kind is called a *scatter diagram*. The small

distribution of scores corresponding to any single rating — as, for example, the 1 score of 10, the 2 scores of 9, and the 1 score of 8 under A — is called an *array of scores*. Similarly, the distribution of rating corresponding to each score — as, for example, the 1, 2, 1 opposite 10 — is called an *array of ratings*.

TABLE 33

A SCATTER DIAGRAM SHOWING THE DISTRIBUTION OF SCORES OF THOSE WHO MADE EACH OF THE FIVE RATINGS

TEACHER'S RATINGS

SCORES	E	D	C	B	A <sup>1</sup>	TOTALS
						4
			≡ 	≡ 		16
		≡ 	≡ ≡	≡ 		24
		≡ 	≡ 			16
						4
Totals	4	16	24	16	4	64

One way of making this table would be to take the A's first and distribute in the A column the scores of those who made A, then take the B's, etc. If this method is used, you must put a check mark after each score as tabulated; otherwise you are sure to miss one or two and you will not know which ones you missed.

Another method is to begin with Pupil No. 1 and take each pupil in order, to find first the appropriate column according to the pupil's rating and then the square in it corresponding

<sup>1</sup> The reason for putting the A on the right will be apparent later.

to his score, and put a mark in that square. In this way you are less likely to miss one, but you must be very careful to get the right square.

Having distributed the scores as shown in Table 33, we see that four pupils had ratings of A and that one made a score of 10, two made 9, and one 8. The mean (average) of these four scores is 9.

*Problem:* What is the mean of the scores of those who got B? *Solution:* Let us remember the short method learned in Chapter I.

ACTUAL SCORES	AUXILIARY SCALE (SUBSTITUTED)	FREQUENCY	PRODUCT
10	4	2	8
9	3	6	18
8	2	6	12
7	1	2	2
		16	40

$$40 \div 16 = 2.5$$

Auxiliary scale value 2.5 = Actual score of 8.5

The mean of the scores of those who got a rating of B, therefore, is 8.5 points.

EXERCISE 35. Find the mean score of those who made C, D, and E, separately, using the short method. Begin with E if you wish.

If your work in the above exercise is correct, you find the means of the arrays of scores corresponding to the five ratings to be as follows:

Rating . . .	E	D	C	B	A
Mean score . .	7	7.5	8	8.5	9

From these figures we see that there is an appreciable tendency for those who made A to get higher scores than those who made B, etc., and the lower the teacher's rating the lower the central tendency of scores. There is, then, some correlation between scores and ratings in this case.

The question arises at this point as to how much rise in the mean scores from array to array might we expect if there were perfect correlation.

**Perfect correlation.** Let us suppose that all the pupils who made A got scores of 10, all who made B got scores of 9, etc., so that our table with the count marks added would appear as shown in Table 34. In that case we should say that there was *perfect correlation* between scores and teacher's ratings.

TABLE 34

SHOWING PERFECT CORRELATION BETWEEN SCORES AND TEACHER'S RATINGS

TEACHER'S RATINGS

SCORES		E	D	C	B	A	TOTALS
	10					4	4
	9				16		16
	8			24			24
	7		16				16
	6	4					4
	Totals	4	16	24	16	4	64

Perfect correlation between two variables denotes the maximum tendency for a high value in one to be accompanied by a high value in the other and vice versa.<sup>1</sup>

In this case the mean score of those who made each rating would be, of course, as follows :

Rating . . . .	E	D	C	B	A
Mean score . .	6	7	8	9	10

<sup>1</sup> The present discussion refers only to what is called positive correlation.

These mean scores show greater increase from low to high. Those obtained in the first case ranged only from 7 to 9, whereas these range from 6 to 10. From this it would appear that the higher the correlation between scores and ratings, the more marked will be the rise in the means of the arrays of scores. This of course is exactly what we should expect, and indeed if there were no rise at all in the means, we should say that there was no correlation between scores and ratings.

At this point the question naturally arises as to whether we may not measure the degree of correlation — the amount of the tendency for high ratings to be accompanied by high scores — in terms of the amount of rise in the means of the arrays of scores for the various ratings. The answer is yes. This is best shown graphically.

Let us first find the line of relation between scores and ratings. In order to do this we must give some numerical values to the ratings A, B, C, D, and E, so that we may find a measure of the variability of the distribution of ratings.

Let us assume that the ratings A, B, C, D, and E represent degrees of ability in arithmetic equidistant upon the scale of merit. We may assign numerical values to the ratings as follows:

Rating	.	.	.	.	.	.	.	E	D	C	B	A
Assigned numerical value	.	.	.	.	.	.	.	1	2	3	4	5

The variability of scores and of the ratings according to their numerical values is the same, since the distributions are identical. Therefore each unit of score corresponds to one unit of rating and the relation line is as shown in Figure 48.

Another line, called the line of means, has been drawn through the point on each vertical line representing the mean of the distribution of scores corresponding to that rating.

Note that while the score corresponding to a rating of A is 10, the mean of the scores corresponding to this rating is

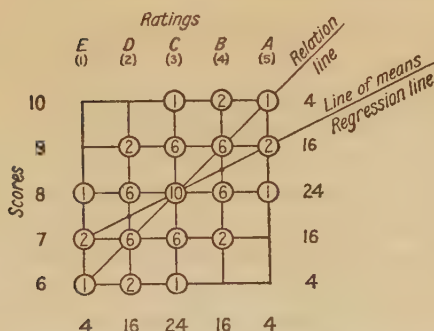


FIG. 48. Showing the relation between a regression line and a relation line.

only 9 (*one half* the way from 8 — the mean of all the scores). Similarly, while the score corresponding to a rating of B is 9, the mean of the scores corresponding to this rating is only 8.5 (*one half* the way from 8). In each case the mean score is only *one half* the way from the mean of all the scores to the score which corresponds to the rating.

Looking at it from another point of view, as we go from one rating to the next the corresponding scores as indicated by the relation line increase by 1 point, whereas the mean scores increase by only .5 of a point. In other words, the line of means is only *one half* as steep as the line of relation. It is said to *regress* toward the horizontal and for that reason is called a *regression line*.

**Coefficient of correlation.** As was suggested, the steeper the line of means the higher the correlation; in the case of perfect correlation the line of means coincides with the line of relation, and in the case of no correlation the line of means is horizontal. Now, since the tendency toward correlation between scores and ratings in the above case was such that the line of means was only half as steep as the line of relation, we say that the scores and ratings are correlated to the extent of  $\frac{50}{100}$ , or .50. Amount of correlation is usually

expressed as a decimal, and the decimal is called the *coefficient of correlation*.

The means of the scores accompanying the various ratings do not always lie in a straight line, as they do in our artificial correlation table. In that case a straight line is supposed to be drawn so as to pass most nearly through the means.

As a coefficient of correlation is usually calculated, no lines are actually drawn but the coefficient is the same as if found in the way described above.

If the correlation is perfect, the line of means is  $\frac{100}{100}$  as steep as the line of relation and the coefficient of correlation is, therefore, 1.00. If there is no correlation and hence no slant at all in the line of means, then the coefficient is .00.

In Figure 49 are shown various ways in which scores and ratings might conceivably be correlated. The coefficients range from .00 to 1.00. These, again, are artificial correlation tables.

**EXERCISE 36.** Convert the second and fourth diagrams in Figure 49 into the form of that shown in Figure 48 — that is, with the numbers on the intersections of the lines. Find the mean of each vertical distribution of scores, plot a point on the vertical line to show the exact location of the mean, and draw a line of means (regression line) similar to that in Figure 48. The line of relation is the same for all. Compare the slant of the regression line with the coefficient of correlation, as was done at the foot of page 181.

**Significance of the coefficient of correlation.** As stated above, the coefficient of correlation is not ordinarily found by

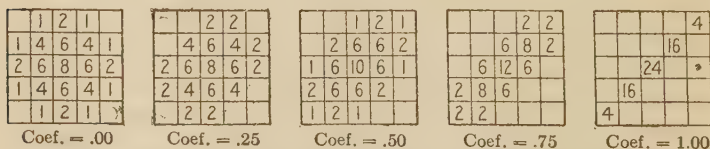


FIG. 49. Showing the general appearance of the scatter diagrams of five typical correlations.

drawing a regression line; the method of finding a coefficient of correlation is explained in full below. The above discussion, however, is calculated to show the meaning and significance of a coefficient of correlation. It is a measure of the degree to which high values of one variable (i.e., ratings) tend to be accompanied by high values of another variable (i.e., scores). It is a measure of the amount of the increase in mean values of one variable (i.e., scores) associated with successive values of the other variable (i.e., ratings) in comparison with the maximum possible increase in such mean values. In other words, it is a measure of the steepness of the line of means (regression line) in comparison with the line of relation.

**A coefficient of correlation in relation to causation.** There is an important significance of a coefficient of correlation which we have not discussed. It is the inference that can be drawn from a coefficient of correlation as to the causal connection between two variables. For example, if there is a correlation of .50 between pupils' scores in a test and teacher's ratings, what can be said about the causes operating to make the teacher's ratings agree with the test scores? This can best be illustrated by means of the flipping of coins.

Let us suppose that we flipped four coins and recorded the number of heads and then left two lying, flipped the other two and recorded the number of heads in all four coins again. Let us suppose that we did this 64 times, flipping first four, then two. According to the law of chance the most probable distribution of heads in the first 64 throws would be as follows:

Number of heads	.	.	.	0	1	2	3	4
Frequency	.	.	.	4	16	24	16	4

Now, when the number of heads is four all the coins are, of course, heads, and if two are left lying these must be heads;

so the number of heads in the second count cannot be less than 2 and may be 2, 3, or 4. By the law of chance the numbers of heads in the second throw will most probably be as follows :

4 heads . .	1
3 heads . .	2
2 heads . .	1

Similarly, according to the law of chance the most probable distribution of heads in the second counts of the 16 cases in which the number of heads by the first count was three will most probably be :

4 heads . .	2
3 heads . .	6
2 heads . .	6
1 head . .	2

Bringing these two and the remaining three cases together, the most probable distributions of numbers of heads in the second throws (two coins selected at random always left lying) will be as follows :

TABLE 35  
NUMBER OF HEADS IN FIRST THROW

NUMBER OF HEADS IN SECOND THROW	0	1	2	3	4
4			1	2	1
3		2	6	6	2
2	1	6	10	6	1
1	2	6	6	2	
0	1	2	1		

We recognize this immediately as our hypothetical case of a correlation of .50.

Now if we tried the same experiment but left only one coin lying, the most probable distributions of numbers of heads on second throw would be just that shown in Figure 49 as yielding a coefficient of correlation of .25. And if 3 out of the 4 are left lying, a scatter diagram yielding a coefficient of .75 will tend to occur.

*A coefficient of correlation, then, may be thought of as the decimal fraction which tells what proportion of the causes affecting the magnitudes of two variables are common to both variables.*

That is, if 10 causes operated to affect a pupil's rating by the teacher, such as mental ability, regularity of attendance, amount of previous instruction, application, personal like or dislike by teacher, health, etc., and 10 similar causes operated to affect the pupil's score in the test, and if these were of equal potency, then if 7 of the 10 causes affecting a pupil's rating also affects his score, the coefficient of correlation between the ratings of scores would be .70. If the correlation coefficient were .80, 8 out of 10 causes are common to both variables.<sup>1</sup>

### QUESTION

Can you think of some of the causes that might be common to ability in arithmetic and ability in language so as to produce a correlation between these abilities even among pupils of the same age and grade?

<sup>1</sup> If there were 10 causes operating in one case and only 8 in the other and 6, say, were common, this would mean of course that  $\frac{6}{10}$ , or 60 per cent, of the causes in one case were the same as  $\frac{6}{8}$ , or 75 per cent, of the causes in the other. In that case the correlation would be  $\sqrt{.60 \times .75}$ , or about .67. All this discussion assumes the causes to be independent of one another (i.e., themselves uncorrelated).

## CHAPTER SIXTEEN

### THE CALCULATION OF A COEFFICIENT OF CORRELATION

THE first method to be used in the finding of a coefficient of correlation was that described above; namely, to find the relative slant of the regression line. The method and the term *regression line* were first used by Sir Francis Galton in dealing with relationships in the biological sciences.

Later Professor Karl Pearson devised a mathematical means of finding the same coefficient (i.e., a method that did not require the finding of a regression line). The method suggested by Pearson is called the *product-moment method*.

The product-moment method is not the best method of finding the coefficient of correlation. A better method is described in the latter part of this chapter. However, this better method is not well known and the product-moment method is in very common use. For that reason, and because of its historic significance, the product-moment method is described briefly herein. Those not interested may skip to the description of the difference method, which is recommended for general use.

**The product-moment method.** Let us suppose that we have a scatter diagram of the scores and teacher's ratings of a group of 64 pupils, as shown in Figure 50. This is an artificial scatter diagram made over-simple for the sake of ease of computation.

The first step is to find the sum of the frequencies in each row and in each column. These are shown in the rectangles at the right and bottom of the large square. Next, these sums in each rectangle are to be totaled to find  $N$ , the number of cases in all. The two totals should be the same, of course.

The next step is to assign an arbitrary auxiliary scale, called the  $x$  scale, to the distribution of ratings having the value 0 opposite the compartment containing the mean of

		Rating						
		E	D	C	B	A		
		(1)	(2)	(3)	(4)	(5)	Sums	y
Score	10				2	2	→ 4	+2
	9			6	8	2	→ 16	+1
	8		6	12	6		→ 24	0
	7	2	8	6			→ 16	-1
	6	2	2				→ 4	-2
Sums		4	16	24	16	4	N = 64	
x		-2	-1	0	+1	+2		

FIG. 50. Showing the calculation of a coefficient of correlation.

the distribution to let the other compartments be given values of + 1, + 2, etc., to the right and - 1, - 2, etc., to the left. And similarly a y scale is assigned to the distribution of scores.

The means of the distributions may be found by any of the various methods described in previous chapters. In our hypothetical case the mean of the scores is exactly 8 and the mean of the ratings is exactly 3. If the mean of the scores had been any value between 7½ and 8½, the assignment of y values to the distribution of scores would have been the same. That is, fractions are usually disregarded in assigning these x and y scale values.

When the mean of either of the two variables contains a fraction, thus to disregard it causes a slight error to be introduced into the coefficient of correlation, but this may be corrected by auxiliary calculation. For purposes of our illustration, however, this correction will be disregarded.

To find  $\Sigma x^2$  (See page 188.)

$$\begin{aligned} 4 \times (+2)^2 &= 16 \\ 16 \times (+1)^2 &= 16 \\ 24 \times 0^2 &= 0 \\ 16 \times (-1)^2 &= 16 \\ 4 \times (-2)^2 &= 16 \\ \Sigma x^2 &= 64 \\ \text{Also } \Sigma y^2 &= 64 \end{aligned}$$

To find  $\Sigma xy$  (See page 188.)

FREQ.	x	y	xy
2	(+1)	(+2)	= +4
2	(+2)	(+2)	= +8
8	(+1)	(+1)	= +8
2	(+2)	(+1)	= +4
2	(-2)	(-1)	= +4
8	(-1)	(-1)	= +8
2	(-2)	(-2)	= +8
2	(-1)	(-2)	= +4
$\Sigma xy =$			48
$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$			$\frac{48}{\sqrt{64 \times 64}}$
$r =$			$\frac{48}{64}$
$r =$			.75

The next step is to find  $\Sigma x^2$ , the sum of the squares of the values of  $x$  in the distribution of ratings. Thus, there are four cases in which  $x = +2$ . The sum of the  $x^2$ 's in these four cases is 16. There are 16 cases in which  $x = +1$ , 24 cases in which  $x = 0$ , etc. The sum of all the  $x^2$ 's is 64. Since the distribution of  $y$  values is exactly the same as the distribution of  $x$  values,  $\Sigma y^2$  is also 64.

The next step is to find  $\Sigma xy$ , the sum of all the products of  $x$  and  $y$ . Thus, there are 2 cases in which  $x = +1$  (Rating = B) and  $y = +2$  (Score = 10). In each of these two cases  $xy = (+1) \times (+2)$  and the sum of the 2  $xy$ 's is  $2 \times (+1) \times (+2) = 4$ . There were 2 cases in which  $x = +2$  and  $y = +2$ , etc.

Note that in all cases in the column of cells corresponding to a rating of C,  $x = 0$ ; therefore  $xy = 0$ ; therefore these cases may be disregarded, since they would in no way affect the sum of all the  $xy$  products. Similarly, in all cases in the row of cells corresponding to a score of 8,  $y = 0$ ; therefore those cases also may be disregarded. There are, therefore, in this case only 8 cells that contain frequencies that need be taken into account. The sum of the  $xy$  products ( $\Sigma xy$ ) in these 8 cells is 48.

All these  $xy$  products happened to be positive. If there had been some frequencies in the lower right-hand corner of the chart,  $x$  would have been positive and  $y$  negative and the  $xy$  products negative (for these cases). For cases in the upper left-hand corner,  $x$  would have been negative and  $y$  positive; so the  $xy$  products of these cases also would have been negative. The value  $\Sigma xy$  is the *algebraic sum* of all the  $xy$  values positive and negative. It is found by subtracting the sum of the negative  $xy$ 's from the sum of the positive  $xy$ 's.

The coefficient of correlation is found by dividing  $\Sigma xy$  by the square root of the product of  $\Sigma x^2$  and  $\Sigma y^2$ . That

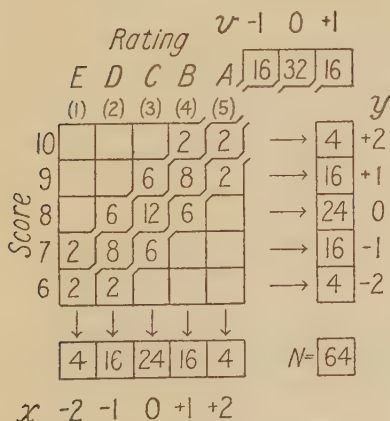
is, by the product-moment formula for a coefficient of correlation (represented by  $r$ ),

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} \quad (\text{Formula 4, Product-moment formula})$$

For the hypothetical case that we have used as an illustration the coefficient of correlation is seen to be .75.

**The difference method.** Considering that to find  $\Sigma xy$  it is necessary to find the sum of the  $xy$  values for each cell separately, it is a little tedious to find a coefficient of correlation by the product-moment method. There is a more convenient method called the *difference method*. This is recommended for general use.

As an illustration of the method let us use the same hypothetical scatter diagram as was used to illustrate the product-moment method. This is shown in Figure 51.



$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma v^2}{2\sqrt{\Sigma x^2 \Sigma y^2}}$$

To find  $\Sigma y^2$

$$\begin{aligned} 4 \times 2^2 &= 16 \\ 16 \times 1^2 &= 16 \\ 24 \times 0^2 &= 0 \\ 16 \times (-1)^2 &= 16 \\ 4 \times (-2)^2 &= 16 \\ \Sigma y^2 &= 64 \\ \text{Also } \Sigma x^2 &= 64 \\ 16 \times 1^2 &= 16 \\ 32 \times 0^2 &= 0 \\ 16 \times (-1)^2 &= 16 \\ \Sigma v^2 &= 32 \end{aligned}$$

$$r = \frac{64 + 64 - 32}{2\sqrt{64 \times 64}}$$

$$\begin{aligned} r &= \frac{96}{128} \\ r &= .75 \end{aligned}$$

FIG. 51. Illustrating the difference method of calculating a coefficient of correlation.

The steps in the general procedure are as follows :

1. *Find the distribution of ratings.* This is done by adding the frequencies in the several columns of squares. The distribution is shown in the rectangle at the foot of the figure. The variable that is plotted in the horizontal direction is called the  $X$  variable. The ratings in this case then constitute the  $X$  variable.

2. *Find the distribution of scores.* This is done by adding the frequencies in the rows of squares. The distribution is shown in the rectangle at the right. The variable plotted in the vertical direction is called the  $Y$  variable. The scores in this case then constitute the  $Y$  variable.

3. *Find the distribution of diagonal sums.* This is shown in the three squares in the upper corner of the figure. It is found by adding the frequencies along each diagonal. This distribution is called the  $V$  distribution. When measured from their own means the values of  $X$ ,  $Y$ , and  $V$  are called respectively  $x$ ,  $y$ , and  $v$ . In the distribution of values of the  $X$  variable there are 4 cases in which  $x = -2$ , 16 cases in which  $x = -1$ , etc.

4. Find the sum of the squares, the values of  $x$ , of  $y$ , and of  $v$ . The sum of the squares of the  $x$  values is denoted by the symbol  $\Sigma x^2$ , and the same for  $\Sigma y^2$  and  $\Sigma v^2$ . The values of  $\Sigma x^2$ ,  $\Sigma y^2$ , and  $\Sigma v^2$  are found as shown in the calculations.

5. Substitute these values of  $\Sigma x^2$ ,  $\Sigma y^2$ , and  $\Sigma v^2$  in the formula,

$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma v^2}{2 \sqrt{\Sigma x^2 \Sigma y^2}}, \quad (\text{Formula 5, Difference formula})$$

in which  $r$  stands for the coefficient of correlation. This substitution and the calculation of the value of  $r$  are shown in Figure 51.

The difficulty with fractions. You will realize, of course, that the scatter diagram used above as an illustration is a

very simple and artificial one with perfectly normal distributions, of which the means came out exactly "even." Ordinarily, of course, the mean of a distribution does not come out even and we must choose between taking the nearest whole number and using a fraction in finding the sums of the squares of the deviations of values from the means, such as  $\Sigma x^2$ .

If the nearest whole number is taken, a certain error is introduced into the calculation which should be avoided. On the other hand, you can see that if the mean of a distribution came out 27.41, for example, it would be extremely tedious to square deviations of .41, 1.41, 2.41, etc., up to 10.41, perhaps, and the same for  $-.59$ ,  $-1.59$ , etc., and add those squares.

**The use of an assumed mean.** Fortunately, however, there is a method of finding the value of  $\Sigma x^2$  without using fractions, that gives the same value as if the exact mean had been used. This method is to take an assumed mean — say the whole number below the true mean — and find the sum of the squares of the deviations of scores from this assumed mean (call this  $\Sigma X^2$ ), and then apply a correction which is not difficult to calculate. The correction is the square of the sum of the deviations themselves, divided by the number of cases. Expressed in a formula:<sup>1</sup>

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{N}. \quad (\text{Formula 6})$$

This same correction can be applied, of course, to all three values  $\Sigma x^2$ ,  $\Sigma y^2$ , and  $\Sigma v^2$  that are needed in the calculation of the coefficient of correlation.

For purposes of calculating a coefficient of correlation it is not necessary to deal with the actual score values themselves.

<sup>1</sup> For the derivation of this formula see L. L. Thurstone, "A Method of Calculating the Pearson Correlation Coefficient without the Use of Deviations," *Psychological Bulletin*, Vol. 14 (1917), pages 27-32.

That is, we may substitute auxiliary scale values in the same way as we did for finding a mean by the "short method" in Chapter II. When referring to the calculation of a coefficient of correlation, therefore,  $x$  and  $y$  refer to deviations from assumed mean in terms of the auxiliary scales of which the value 0 is the assumed mean of the distribution. (See Figures 51 and 52.)

**The Otis Correlation Chart.** When this correction is introduced into the formula

$$r = \frac{\Sigma x^2 + \Sigma y^2 - \Sigma v^2}{2 \sqrt{\Sigma x^2 \Sigma y^2}},$$

the resulting formula is rather complicated; and while the calculation of the coefficient is very easy, the steps are many and one is likely to be overlooked. For that reason, it has seemed desirable to draw up a chart for use in finding a coefficient of correlation in which a "job analysis" has been made of the various steps necessary in the calculation and then laid out in order so that no attention need be paid to the formula while making the calculation.

The Otis Correlation Chart<sup>1</sup> fulfills this need. (See Figure 52.) It is so arranged that any one can find a coefficient of correlation by its use without even knowing what formula is being used. All that is necessary is a careful reading of the directions and accurate arithmetic computation. The reader should note that not all of the calculations provided for the chart are necessary to find a coefficient of correlation. The chart also contains provision for finding the probable error of the coefficient due to the limited number of cases, and directions for finding the true means and the standard deviation

<sup>1</sup> The Otis Correlation Chart is published by World Book Company, Yonkers-on-Hudson, New York. It may be bought in packages of 25, with full directions for use and sample chart as shown in Figure 52. Price \$1.25 per package. The charts are  $8\frac{1}{2}$  by 11 inches, with  $\frac{1}{4}$ -inch squares. Each chart contains a Table of Products and a Table of Squares on the back.

of the variables that are being correlated. It is not necessary to find these unless they are specially needed, and often they may be omitted.

It is not deemed to be within the scope of this elementary book to discuss further the theory underlying the formula used in the Otis Correlation Chart. Let it suffice to say that the formula is the same as that used in the illustrative example in Figure 51, with the corrections applied which do away with fractions but yet retain the full accuracy of the coefficient. The advantages of the chart over the product-moment method are further set forth in an article entitled "The Otis Correlation Chart," in the *Journal of Educational Research*, December, 1923, Vol. VIII, No. 5, page 440.

In order not to confuse those who are not familiar with the symbols  $\Sigma x^2$ ,  $\Sigma y^2$ , etc., these values have been replaced in the formula and in the computations by single letters. The correlation formula as used is given in Figure 52. The symbol  $r_{xy}$  stands for the coefficient of correlation between two variables  $x$  and  $y$ .

**How to use the Correlation Chart.** While reading these directions refer to Figure 52.

The first main operation is the plotting of the paired values of the two variables to be correlated. Before doing this, it is necessary (1) to call one the  $x$  variable and the other the  $y$  variable, (2) to choose the interval in each for grouping, and (3) to label the plot for locating the distributions.

1. *Designating the variables.* If one variable is thought of as depending on the other, the independent variable is usually called the  $x$  variable and the dependent variable the  $y$  variable. This is optional. The choice should be determined chiefly by the ease of plotting. Having designated the variables,  $x$  and  $y$ , enter their names at the top of the sheet.

2. *Choosing the interval.* If the number of units in the range of a variable is not over 19 (as, for example, if the values

range only from 15 to 33), no grouping is necessary. If the range is more than 19 points, however, values must be grouped. It is customary to group by 2's, 5's, or 10's, thus: 0-1, 2-3, 4-5, etc., or 0-4, 5-9, 10-14, etc., or 0-9, 10-19, 20-29, etc. These are called class intervals, or better, class ranges.

It is suggested that the values be grouped if possible so that the number of groups will be between 10 and 15.

3. *Labeling the plot to locate the distributions.* Having chosen the interval for grouping, label the plot (large square) at the bottom and left side by writing the class ranges in order. This should be done in such a way that the distributions of  $x$  and  $y$  values will be centered as nearly as possible on the plot. It is planned that the 0 points in the  $X$  and  $Y$  scales at the bottom and right side of the plot shall be a little off the center of the distributions.

In order to center the distributions, it is convenient to determine the lowest and highest values of each variable and to find the number of class intervals covered by the range from lowest to highest. This range for each variable may then be fitted to the plot so that it will be approximately centered. Always label the plots so that the  $X$  values increase to the right and the  $Y$  values upward.

4. *Plotting the points.* Plot a "point" for each pair of values to be correlated, placing it in the row of squares opposite the interval within which the  $y$  value falls and in the column above the interval within which the  $x$  value falls.

In plotting the "points" it is customary not to make dots but tallies. These are more easily counted.

*Use of T-square and triangle.* To avoid fatigue and frequent errors in plotting, it is very desirable to use an auxiliary strip of paper, containing the intervals of one variable, which can be moved over into the plot when locating the points. Thus, suppose the  $x$  values are to be read first. Copy the

# OTIS CORRELATION CHART

By Arthur S. Otis, Ph.D.  
Author of the Otis Group Intelligence Scale

Correlation between Higher Exam Form A (x) and Higher Exam Form B (y)  $r_{xy} = .917 \pm .009$

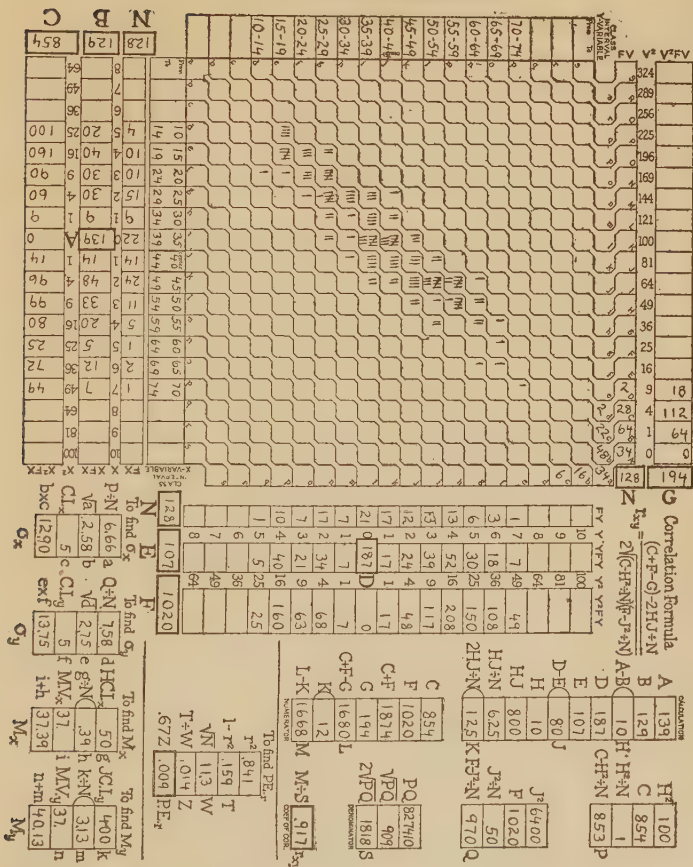


FIG. 52. Illustrating the calculation of a coefficient of correlation by means of the Otis Correlation Chart. (Actual size of chart, 8 by 11 inches.)

class ranges of the  $y$  variable on to a strip of paper. Move the strip so that it is adjacent to the column in which the point must fall; then locate the square within the column by means of the strip. In this way the danger of error is greatly reduced.

As an aid in keeping the strip properly adjusted, it may be mounted on a triangle or square piece of cardboard and slid back and forth along a T-square or other straight-edge, fastened in position, horizontally. This method is habitually used by statisticians.

5. *Finding the distributions.* Add the tallies in three directions: horizontally, vertically, and diagonally.

First count the tallies in each row and place the sum in the column headed  $FY$  (Frequency of  $Y$ ; i.e., the number of values of  $Y$  in each interval). Total these and write the sum in box  $N$ .

Next, turn the paper counter-clockwise a quarter turn and count the tallies in the columns. Place the sums in the column headed  $FX$  (Frequency of  $X$ ). Total these and write the sum in box  $N$ . This sum should check with the sum in the other box  $N$ . If it does not, repeat the adding of rows and columns.

Next, turn the paper back to the original position and one-eighth turn farther. The diagonals will then be horizontal. Each will begin with a small letter at the left. Count the tallies in each diagonal and write the sum in the square having the same letter just outside the heavy line either at the top or right side of the plot.

Next, the distribution of diagonal sums at the right side of the plot is to be combined with the distribution at the top. Copy in the  $A$  square the sum which is in the  $a$  square. Write in the  $B$  square the sum of the numbers in the two  $b$  squares. Write in the  $C$  square the sum of the numbers in the two  $c$  squares, etc. (When thoroughly familiar with these steps, it

is permissible to add all the tallies in both  $b$  diagonals in a single operation and place the sum immediately in the large  $B$  square, etc.)

6. *To find G.* Add the numbers in the  $FV$  column just filled and place the sum in box  $N$ . This sum should check with the sums in the other  $N$  boxes. If it does not, add the diagonals again, looking for stray tallies.

Next, multiply each number in the  $FV$  column by the adjacent number in the  $V^2$  column.<sup>1</sup> Write the products in the  $V^2FV$  column.<sup>2</sup>

Next, add the numbers in the  $V^2FV$  column and write the sum in the  $G$  box.

7. *To find D, E, and F.* Multiply each number in the  $FY$  column by the adjacent number printed in the  $Y$  column. Write the product in the  $YFY$  column. Do not write a product in the  $D$  box; this would in any case be 0. Add the numbers in the  $YFY$  column above the  $D$  box and write the sum in the  $D$  box. Add the numbers in the  $YFY$  column below the  $D$  box and write the sum in the  $E$  box.

Next, multiply each number in the  $YFY$  column by the number adjacent to it in the  $Y$  column (not  $Y^2$ ) and write the product in the  $Y^2FY$  column. Each number in the  $FY$  column will then have been multiplied by  $Y$  twice, and is therefore  $Y^2$  times  $FY$  in the  $Y^2FY$  column. The values of  $Y^2$  also are given, in order that  $Y^2FY$  can be obtained by multiplying  $FY$  by  $Y^2$  if this is easier than to multiply  $YFY$  by  $Y$ . The result is, of course, the same. There will be nothing to put in the space to the right of  $D$ . Any number times  $0^2$  equals 0.

Next, add the numbers in the  $Y^2FY$  column and write the sum in the  $F$  box.

<sup>1</sup> See page 199 regarding use of a table of products to aid in this.

<sup>2</sup>  $V$  stands for the value,  $Y - X$ . If  $Y = 6$  and  $X = 4$ , then  $V = 2$  and  $V^2 = 4$ .  $FV$  means Frequency of  $V$  Values.  $V^2FV$  means the Frequency of  $V$  values multiplied by  $V^2$ .

Table of Products										Table of Squares																													
40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
800	778	756	734	712	690	668	646	624	602	580	558	536	514	492	470	448	426	404	382	360	338	316	294	272	250	228	206	184	162	140	118	96	74	52	30	8	1		
810	788	766	744	722	700	678	656	634	612	590	568	546	524	502	480	458	436	414	392	370	348	326	304	282	260	238	216	194	172	150	128	106	84	62	40	18	1		
820	798	776	754	732	710	688	666	644	622	600	578	556	534	512	490	468	446	424	402	380	358	336	314	292	270	248	226	204	182	160	138	116	94	72	50	28	1		
830	808	786	764	742	720	698	676	654	632	610	588	566	544	522	500	478	456	434	412	390	368	346	324	302	280	258	236	214	192	170	148	126	104	82	60	38	1		
840	818	796	774	752	730	708	686	664	642	620	598	576	554	532	510	488	466	444	422	400	378	356	334	312	290	268	246	224	202	180	158	136	114	92	70	48	1		
850	828	806	784	762	740	718	696	674	652	630	608	586	564	542	520	498	476	454	432	410	388	366	344	322	300	278	256	234	212	190	168	146	124	102	80	58	1		
860	838	816	794	772	750	728	706	684	662	640	618	596	574	552	530	508	486	464	442	420	398	376	354	332	310	288	266	244	222	200	178	156	134	112	90	68	1		
870	848	826	804	782	760	738	716	694	672	650	628	606	584	562	540	518	496	474	452	430	408	386	364	342	320	298	276	254	232	210	188	166	144	122	100	78	1		
880	858	836	814	792	770	748	726	704	682	660	638	616	594	572	550	528	506	484	462	440	418	396	374	352	330	308	286	264	242	220	198	176	154	132	110	88	1		
890	868	846	824	802	780	758	736	714	692	670	648	626	604	582	560	538	516	494	472	450	428	406	384	362	340	318	296	274	252	230	208	186	164	142	120	98	1		
900	878	856	834	812	790	768	746	724	702	680	658	636	614	592	570	548	526	504	482	460	438	416	394	372	350	328	306	284	262	240	218	196	174	152	130	108	1		
910	888	866	844	822	800	778	756	734	712	690	668	646	624	602	580	558	536	514	492	470	448	426	404	382	360	338	316	294	272	250	228	206	184	162	140	118	1		
920	898	876	854	832	810	788	766	744	722	700	678	656	634	612	590	568	546	524	502	480	458	436	414	392	370	348	326	304	282	260	238	216	194	172	150	128	1		
930	908	886	864	842	820	798	776	754	732	710	688	666	644	622	600	578	556	534	512	490	468	446	424	402	380	358	336	314	292	270	248	226	204	182	160	138	1		
940	918	896	874	852	830	808	786	764	742	720	698	676	654	632	610	588	566	544	522	500	478	456	434	412	390	368	346	324	302	280	258	236	214	192	170	148	1		
950	928	906	884	862	840	818	796	774	752	730	708	686	664	642	620	598	576	554	532	510	488	466	444	422	400	378	356	334	312	290	268	246	224	202	180	158	1		
960	938	916	894	872	850	828	806	784	762	740	718	696	674	652	630	608	586	564	542	520	498	476	454	432	410	388	366	344	322	300	278	256	234	212	190	168	1		
970	948	926	904	882	860	838	816	794	772	750	728	706	684	662	640	618	596	574	552	530	508	486	464	442	420	398	376	354	332	310	288	266	244	222	200	178	156	1	
980	958	936	914	892	870	848	826	804	782	760	738	716	694	672	650	628	606	584	562	540	518	496	474	452	430	408	386	364	342	320	298	276	254	232	210	188	166	1	
990	968	946	924	902	880	858	836	814	792	770	748	726	704	682	660	638	616	594	572	550	528	506	484	462	440	418	396	374	352	330	308	286	264	242	220	198	176	154	1
1000	978	956	934	912	890	868	846	824	802	780	758	736	714	692	670	648	626	604	582	560	538	516	494	472	450	428	406	384	362	340	318	296	274	252	230	208	186	164	1

FIG. 53. Table of Products for use with the Otis Correlation Chart. (Actual size, 8 by 10½ inches.)

8. *To find A, B, and C.* Do exactly the same thing with the numbers in the  $FX$  column as was done with the numbers in the  $FY$  column. (Read  $X$  for  $Y$  above and  $A, B, C$  for  $D, E$ , and  $F$ , and the directions will be identical.)

Boxes  $A, B, C, D, E, F$ , and  $G$  will now be filled.

*How to use the Table of Products.* The Table of Products on the back of the correlation chart is provided for simplifying the finding of the products,  $AFX, X^2FX, YFY, Y^2FY$ , and  $V^2FV$ . This is shown in Figure 53. The method of using the Table of Products is as follows:

Bring the right-hand edge of the table up to the  $FX$  column, and adjust it so that the heavy horizontal lines are opposite the  $A$  box. Find the column in the table headed by the number corresponding to the given value of  $FX$ . In that column and row will be found the values of  $AFX$  and  $X^2FX$ . Thus, if the Frequency of  $X$  is 7, and this is to be multiplied by 4 and 16, the products, 28 and 112, will be found in the column of the table headed 7 in the row opposite 4 and 16. These products may be written immediately in the  $AFX$  and  $X^2FX$  columns.

The table works in a similar manner for  $YFY$  and  $Y^2FY$  values and for  $V^2FV$  values. In the latter case, of course, read only the lower number in the table.

To aid in locating the proper numbers in the table, a second copy of the table may be placed over the first so that the upper edge containing the frequency numbers is just below the row of numbers in which the products are found. The column may then be located quickly and accurately by means of the auxiliary scale.

9. *To find M.* Copy values  $A$  and  $B$  from the boxes into the spaces marked  $A$  and  $B$  in the column headed "Calculation." Subtract  $B$  from  $A$ . The difference will be known as  $H$ . (If the distributions have been properly placed,  $A$  will exceed  $B$  and there will be no negative values

in the calculation. If  $B$  exceeds  $A$ , subtract  $A$  from  $B$  and place a minus sign in the semicircle.)

Copy  $D$  and  $E$  and subtract  $E$  from  $D$ . (If  $E$  exceeds  $D$ , place a minus sign in the semicircle.) The difference will be known as  $J$ .

Copy  $H$  below  $J$  and multiply. Next, divide the product by  $N$ . Next, multiply by 2. This gives  $K$ .<sup>1</sup> If *either*  $H$  or  $J$  alone is negative,  $K$  is negative; otherwise — if both  $H$  and  $J$  are positive or both negative —  $K$  is positive. As soon as  $H$  and  $J$  are found, if one (but only one) is negative, enter the minus sign of  $K$  immediately in both semicircles below.

Copy  $C$  and  $F$  and add. Copy  $G$  and subtract.<sup>2</sup> Copy  $K$  and subtract.<sup>3</sup> (If  $K$  is negative, add the value of  $K$  to  $L$ .) This gives  $M$ , the numerator of the fraction in the formula.

10. *To find S.* Multiply  $H$  by itself<sup>4</sup> and enter  $H$  at the top of the next column. Copy  $C$ . Divide  $H^2$  by  $N$  and subtract from  $C$ . This gives  $P$ .

Square  $J$ . Copy  $F$ . Divide  $J^2$  by  $N$  and subtract from  $F$ . This gives  $Q$ . ( $H^2$  and  $K^2$  are always positive.)

Multiply  $P$  by  $Q$ . Take the square root of the product and multiply by 2. This gives  $S$ , the denominator.

11. *Short cut to find S.* If  $P$  and  $Q$  do not differ by more than 10 per cent, the value of  $P + Q$  may be substituted for  $2\sqrt{PQ}$  where extreme accuracy is not required, for  $P + Q$  in this case will not differ from  $2\sqrt{PQ}$  by more than about  $\frac{1}{10}$  of 1 per cent.

<sup>1</sup> When  $N$  is even, a short method of finding  $K$  is to divide  $HJ$  by one-half  $N$ .

<sup>2</sup> *Negative correlation.* It is possible for  $M$  to be negative either because  $G$  is greater than  $C + F$  or  $K$  is greater than  $L$ . In that case the coefficient will be negative. (See next page.)

<sup>3</sup> If  $L$  is more than 1000, it is sufficient for ordinary purposes to take the nearest whole number as the value of  $K$ .

<sup>4</sup> This may be done by means of the Table of Squares. It is sufficient for ordinary purposes to take the nearest whole number as the value of  $H^2 \div N$  or  $J^2 \div N$ , when these are less than  $\frac{1}{1000}$  of  $C$  or  $F$ .

12. To find  $r$ , the coefficient of correlation. Divide  $M$  by  $S$ . If only the coefficient of correlation is sought, this is as far as the calculations need to be carried.

Directions for the use of the portions of the chart labeled "To find P.E.<sub>r</sub>," "To find  $\sigma_x$ ," etc., are given in the "Directions" accompanying the charts.

**Negative correlation.** So far we have discussed only the tendency for high values of one variable to be associated with high values of another. This tendency may vary from no tendency at all (no correlation) up to a complete tendency (perfect correlation). The coefficients of (positive) correlation within these extremes vary as we have said, from .00 to 1.00.

It is possible, however, for there to be a tendency for high values of one variable to be accompanied by low values of another. Thus, we find a tendency for high IQ's to be accompanied by low accomplishment ratios. In that case we say there is negative correlation. Negative correlation may range theoretically all the way from no tendency up to a complete tendency or perfect negative correlation. Coefficients of negative correlation range from .00 to  $-1.00$ . Actual instances of negative correlation are very rare.

**EXERCISE 37.** Plot on the Otis Correlation Chart the pairs of scores and ratings given in Table 32 and find the coefficient of correlation according to directions. It should come out .50, of course.

**EXERCISE 38.** By means of an Otis Correlation Chart find the correlation between Part I of the Otis Classification Test (Achievement Test) and the Stanford Achievement Test for the 26 cases given in Table 36. Let the scores in the Otis Achievement Test constitute the  $x$  variable and those in the Stanford Achievement Test the  $y$  variable. Group the scores by 5's in each case. Let the interval 50-54 in the  $x$  variable and the interval 60-64 in the  $y$  variable come in the "center" compartments. This will enable you to compare your work with Figure 54 as a check.



TABLE 36

SHOWING THE SCORES OF 26 PUPILS IN GRADE 8 OF MAPLETON JUNIOR HIGH SCHOOL<sup>1</sup> IN PART I OF THE OTIS CLASSIFICATION TEST (ACHIEVEMENT TEST) AND THE STANFORD ACHIEVEMENT TEST

PUPIL	OTIS ACHIEVEMENT TEST	STANFORD ACHIEVEMENT TEST	PUPIL	OTIS ACHIEVEMENT TEST	STANFORD ACHIEVEMENT TEST
1	89	88.5	14	34	46.5
2	71	79.8	15	14	36.3
3	59	77.2	16	82	83.7
4	54	71.8	17	53	77.2
5	66	70.0	18	57	75.6
6	63	69.7	19	51	66.7
7	54	67.3	20	46	58.3
8	46	62.8	21	43	57.0
9	49	61.7	22	51	56.8
10	44	59.8	23	28	54.0
11	39	59.7	24	38	42.9
12	38	55.1	25	23	42.1
13	40	54.9	26	29	39.5

EXERCISE 39. By means of an Otis Correlation Chart find the correlation between the Higher Examination and the Terman Group Test of Mental Ability for the 108 cases given in Table 37. Let the scores in the Higher Examination constitute the  $x$  variable, with the interval 50-54 in the "center" compartment. Let the interval 150-159 in the Terman Group Test come in the "center" compartment. You can then check your work by referring to Figure 55.

EXERCISE 40. Find the coefficient of correlation between scores in the Terman Group Test and "Scholarship Average" as reported for the 108 cases in Table 37.

EXERCISE 41. Find the correlation between the Higher Examination and Scholarship Average. (See Table 37.)

<sup>1</sup> Data furnished by Elmer H. Webber, Superintendent of Schools, Mapleton, Maine.

## OTIS CORRELATION CHART

By *Arthur S. Otis, Ph.D.*

Author of the Otis Group Intelligence Scale

Correlation between Highs Examination (x) and Terman Group Test

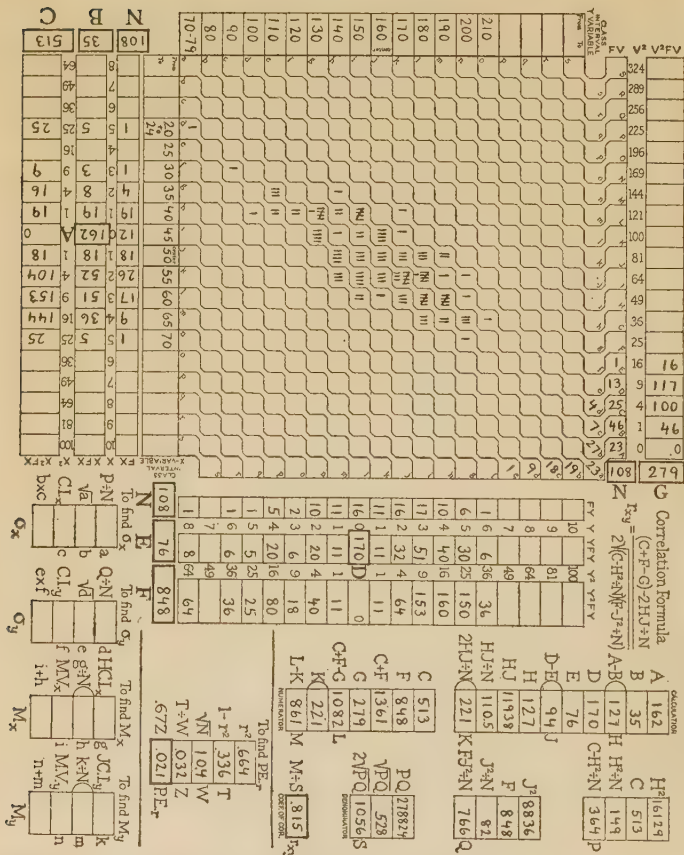
$$r_{xy} = \frac{.82}{.97} = .845$$


FIG. 55. The correlation between the Higher Examination and the Terman Group Test of Mental Ability.

TABLE 37

SHOWING THE SCORES OF 108 FRESHMEN OF RIPON COLLEGE, RIPON, WISCONSIN, IN THE HIGHER EXAMINATION, AND IN THE TERMAN GROUP TEST OF MENTAL ABILITY <sup>1</sup>

CASE	TERMAN GROUP TEST	HIGHER EXAM.	SCHOLAR- SHIP AV.	CASE	TERMAN GROUP TEST	HIGHER EXAM.	SCHOLAR- SHIP AV.	CASE	TERMAN GROUP TEST	HIGHER EXAM.	SCHOLAR- SHIP AV.
1	191	61	84	37	166	49	87	73	171	48	88
2	204	55	90	38	190	52	90	74	160	47	83
3	151	51	76	39	170	56	79	75	192	62	90
4	119	36	79	40	200	69	90	76	193	59	85
5	90	31	66	41	161	51	73	77	195	51	89
6	173	50	87	42	177	63	85	78	186	63	87
7	154	56	70	43	163	50	73	79	180	53	86
8	153	50	87	44	181	58	86	80	207	69	94
9	159	59	82	45	146	44	81	81	173	55	90
10	188	69	83	46	157	44	82	82	129	42	84
11	161	57	78	47	207	66	90	83	171	60	87
12	153	48	86	48	170	57	83	84	187	55	87
13	165	60	84	49	113	40	75	85	130	49	81
14	183	65	86	50	199	67	88	86	134	43	75
15	189	62	89	51	157	64	89	87	181	62	89
16	141	45	77	52	136	46	84	88	158	40	85
17	146	37	81	53	168	49	82	89	155	58	77
18	187	55	86	54	161	49	80	90	179	43	75
19	144	54	81	55	172	53	92	91	168	55	81
20	144	51	86	56	116	38	79	92	216	66	86
21	151	49	78	57	180	55	85	93	131	41	75
22	149	55	78	58	182	54	91	94	177	53	87
23	158	58	91	59	143	54	85	95	142	57	88
24	173	57	72	60	137	43	73	96	153	43	83
25	205	61	85	61	176	55	89	97	105	43	73
26	154	42	70	62	142	55	84	98	155	60	88
27	74	24	61	63	146	54	85	99	119	39	84
28	188	51	88	64	168	56	83	100	181	59	81
29	154	44	85	65	131	43	86	101	125	40	74
30	175	59	87	66	137	47	84	102	177	59	88
31	175	62	77	67	135	44	60	103	167	50	78
32	111	44	83	68	182	60	91	104	182	59	92
33	185	69	84	69	190	63	90	105	177	57	79
34	199	65	87	70	193	61	87	106	187	60	90
35	190	63	83	71	136	46	91	107	135	42	74
36	201	72	91	72	141	44	80	108	152	50	84

<sup>1</sup> Data furnished by Professor B. P. Heubner of the Department of Education and Psychology, Ripon College, Ripon, Wisconsin.

EXERCISE 42. For further practice in the computation involved, find the coefficient of correlation between the Otis Achievement Test and the Stanford Achievement Test from the same data as used above, but label the plots with each distribution moved up one compartment. That is, let the interval 45-49 in the Otis Achievement Test and interval 55-59 of the Stanford Achievement Test come in the center compartments. This simply moves the whole scatter diagram up one square toward the upper right-hand corner of the chart and makes the calculation different throughout, except that  $M$ ,  $P$ , and  $Q$  will come out the same and, of course, the coefficient of correlation must come out the same. The coefficient of correlation is not affected in any way by the position of the scatter diagram on the chart. This same shifting of distributions can be done also in the other cases, if desired.

EXERCISE 43. For additional practice, find the coefficients of correlation between other pairs of tests from the scores given in Table 38. Nine other combinations can be used. Group N. I. T. and T. G. T. scores by 10's and O. I. E. scores by 5's.

For still further practice see Appendix III.

**Rank methods of computing correlation.** Two formulas have been developed by Spearman for finding a coefficient of correlation between two variables, not from the values themselves of the variables but from the ranks of the individuals in the two distributions.

It is obvious that if two tests are perfectly correlated, the rank order of the pupils according to score in one test will be exactly the same as the rank order of the pupils according to score in the other test — the pupil who ranks highest in one test will rank highest in the other; the one who ranks next to the highest in one will rank next to the highest in the other, and so on. But if the correlation is not perfect, there may be some lack of agreement between the rank orders of the pupils in the two tests. For example, the pupil who scores highest in one test might score second from the highest in the other test, the pupil who scored next to highest in the first test might score highest in the second test, and so on.

TABLE 38

SHOWING THE SCORES OF THE 26 MAPLETON JUNIOR HIGH SCHOOL  
PUPILS IN VARIOUS TESTS

PUPIL	OTIS ACHIEVEMENT TEST	STANFORD ACHIEVEMENT TEST	NATIONAL INTELLIGENCE TEST	TERMAN GROUP TEST	OTIS INTER- MEDIATE EXAMINATION
1	89	88.5	111	108	65
2	71	79.8	141	99	45
3	59	77.2	123	87	54
4	54	71.8	111	41	37
5	66	70.0	111	57	39
6	63	69.7	118	72	48
7	54	67.3	86	49	23
8	46	62.8	121	66	42
9	49	61.7	99	51	35
10	44	59.8	124	53	38
11	39	59.7	81	51	33
12	38	55.1	78	48	22
13	40	54.9	91	48	39
14	34	46.5	81	32	25
15	14	36.3	78	31	20
16	82	83.7	112	107	65
17	53	77.2	118	62	52
18	57	75.6	101	69	39
19	51	66.7	72	35	36
20	46	58.3	89	33	39
21	43	57.0	93	64	43
22	51	56.8	92	45	42
23	28	54.0	57	31	33
24	38	42.9	57	28	18
25	23	42.1	48	23	12
26	29	39.5	79	30	22

Obviously, the higher the correlation is, the less will be the aggregate amount of these differences. Now it is possible by taking account of these differences in rank, to work out by means of an appropriate formula a coefficient of correlation between the two tests.

In finding ranks for this purpose, it is customary to assign the highest score in a distribution the rank of 1, the next highest a rank of 2, etc.<sup>1</sup> If two or more individuals have the same scores, each is assigned the average or median of the ranks of the two or more individuals.

The two rank methods are illustrated by Table 39. In this table are reproduced the same scores in the Otis and Stanford Achievement Tests that were given in Table 36. In the fourth column are given the ranks of the several pupils in the Otis Achievement Test, and in the fifth column the ranks of the same pupils in the Stanford Achievement Test. Pupil No. 2, for example, ranked 3 in each test. There were two scores of 77.2 in the Stanford Achievement Test, and these would have received the ranks of 4 and 5; so each was assigned the rank of  $4\frac{1}{2}$ .

In the next three columns are entered the differences between the ranks of pupils in the two tests. It is necessary to keep the positive and negative differences separate. Pupil No. 3 ranked higher in the Stanford Achievement Test because the rank is represented by a lower number. We may choose arbitrarily to let differences in this direction be called positive. (We could just as well call them negative and the others positive.) The difference between the ranks, 6 and  $4\frac{1}{2}$ , of Pupil No. 3 is  $1\frac{1}{2}$ , and the larger number, 6, is on the left; so the  $1\frac{1}{2}$  is placed to the left of the column of zeros. In the case of Pupil No. 5, the larger number, 8, is on the right; so the difference, 4, is placed to the right of the column of zeros. This plan helps one to get all the positive

<sup>1</sup> In a previous discussion of ranks we gave the lowest individual of a group the rank of 1 in anticipation of the discussion of percentile ranks in which high ranks are represented by high numbers, and if desired the same procedure may be used for finding a coefficient of correlation by either of the two rank methods. It is obvious that the difference between the highest rank and the next to the highest rank is the same, whether it is the difference between ranks of 1 and 2 or 26 and 25 (supposing there are 26 individuals in the group in question).

TABLE 39

SHOWING THE METHOD OF FINDING A COEFFICIENT OF CORRELATION BY  
THE TWO RANK METHODS

PUPIL	OTIS ACHIEVE- MENT TEST	STANFORD ACHIEVE- MENT TEST	OTIS ACH. TEST RANK	ST. ACH. TEST RANK	DIFFERENCE + 0 -	SQUARE OF DIF- FERENCE
1	89	88.5	1	1	0	0
2	71	87.9	3	3	0	0
3	59	77.2	6	4½	1½	2¼
4	54	71.8	8½	7	1½	2¼
5	66	70.0	4	8	4	16
6	63	69.7	5	9	4	16
7	54	67.3	8½	10	1½	2¼
8	46	62.8	14½	12	2½	6¼
9	49	61.7	13	13	0	0
10	44	59.8	16	14	2	4
11	39	59.7	19	15	4	16
12	38	55.1	20½	19	1½	2¼
13	40	54.9	18	20	2	4
14	34	46.5	22	22	0	0
15	14	36.3	26	26	0	0
16	82	83.7	2	2	0	0
17	53	77.2	10	4½	5½	30¼
18	57	75.6	7	6	1	1
19	51	66.7	11½	11	0½	0¼
20	46	58.3	14½	16	1½	2¼
21	43	57.0	17	17	0	0
22	51	56.8	11½	18	6½	42¼
23	28	54.0	24	21	3	9
24	38	42.9	20½	23	2½	6¼
25	23	42.1	25	24	1	1
26	29	39.5	23	25	2	4
					24 24	167.5

$$\Sigma g = 24. \quad R = 1 - \frac{6 \times 24}{26^2 - 1} = 1 - \frac{144}{675} = .79. \quad \text{By Table III}^1 r = .95.$$

$$\Sigma D^2 = 167.5. \quad \rho = 1 - \frac{6 \times 167.5}{26(26^2 - 1)} = 1 - \frac{1005}{17550} = .94. \quad \text{By Table IV}^1 r = .95.$$

<sup>1</sup> In the Appendix.

differences in one column and all the negative differences in the other.

We have next to find the sum of the positive differences ("gains"). This is represented by the symbol  $\Sigma g$ . It is customary to find also the sum of the negative differences as a check. This must come out the same as the sum of the positive differences, for every gain in rank of one pupil in one direction is compensated for by some gain of some other pupil or pupils in the other direction. In this case the sums of the positive and of the negative differences are both 24. Therefore  $\Sigma g = 24$ .

**The Spearman Footrule.** The simpler of the two rank methods makes use of what Spearman calls the "Footrule." A value  $R$  is first found by means of the "Footrule," and the coefficient of correlation  $r$  is then found from  $R$ . By the "Footrule,"

$$R = 1 - \frac{6 \Sigma g}{N^2 - 1}, \quad \text{(Formula 7, Spearman's "Footrule")}$$

In the illustrative case  $\Sigma g$ , the sum of the gains is 24. Substituting this value in the "Footrule" formula,

$$R = 1 - \frac{6 \times 24}{26^2 - 1},$$

$$R = .79.$$

The next step is to find the coefficient of correlation ( $r$ ) from the value  $R$ . This is done by means of Table III,<sup>1</sup> page 295. By the table the value of  $r$  corresponding to the value .79 of  $R$  is found to be .952, which we may call .95.

<sup>1</sup> This table is based on the formula

$$r = 2 \cos \frac{\pi}{3} (1 - R) - 1, \quad \text{(Formula 8)}$$

in which

$$R = 1 - \frac{6 \Sigma g}{N^2 - 1} \text{ and } \frac{\pi}{3} = 60^\circ.$$

**The second rank method.** An alternative method — one that takes a little longer but gives slightly more reliable results — involves the finding of the squares of the differences between ranks, as shown in the last column of Table 39. The sum of the squares of the differences in rank is found and represented by the symbol  $\Sigma D^2$ .

A value,  $\rho$ , is then found from  $\Sigma D^2$  by the formula

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2 - 1)} \quad \text{(Formula 9, Second rank formula)}$$

and the correlation coefficient,  $r$ , is found from  $\rho$  by means of Table IV<sup>1</sup> on page 296.

In our illustrative case (Table 39),  $\Sigma D^2 = 167.5$ . Substituting this value in the formula,

$$\begin{aligned} \rho &= 1 - \frac{6 \times 167.5}{26(26^2 - 1)}, \\ \rho &= .94. \end{aligned}$$

Looking up the value .94 of  $\rho$  in Table IV, we find the corresponding value of  $r$  to be .945<sup>+</sup>, which we may call .95.

**Interpolation.** In our initial calculation of  $R$  and  $\rho$  we expressed the values of these measures in three-place numbers. That is, Table III gives .945<sup>+</sup> as the value of  $r$  corresponding to the value .79 of  $R$ . Strictly speaking, the value .945<sup>+</sup> of  $r$  corresponds to the value .790 of  $R$ . The value of  $r$  was recorded as .79, but it was not necessarily .790. It might have been .791 or .792, for example, in which case the corresponding value of  $r$  might be .946. If we wish to get third-place accuracy in value of  $r$ , we must have at least third-place accuracy in the value of  $R$ . All this applies equally, of course, to the calculation of  $r$  from  $\rho$ .

<sup>1</sup> Table IV is compiled by means of the formula

$$r = 2 \sin \left( \frac{\pi}{6} \rho \right), \quad \text{(Formula 10)}$$

in which

$$\frac{\pi}{6} = 30^\circ.$$

Carrying the calculation in Table 39 to the third place,  $R = .787$ . If we wish, we may find the value of  $r$  from Table III corresponding to the value .787 of  $R$ .

The value .787 is  $\frac{7}{10}$  of the way from .78 to .79. Therefore the corresponding value of  $r$  by Table III is  $\frac{7}{10}$  of the way from .947 to .952. The difference between .947 and .952 is .005;  $\frac{7}{10}$  of .005 is .0035; let us call it .004. Adding .004 to .947, we get .951 as the more precise value of the coefficient of correlation by the "Footrule." This operation is called *interpolation*.

*Problem:* Find the value of the coefficient by the second rank method to the third decimal place by interpolation. Try this yourself before reading the solution. *Solution:* Carrying the calculation of  $\rho$  to the third decimal place in Table 39, we find that  $\rho = .943$ .

The value .943 is  $\frac{3}{10}$  of the way from .94 to .95. Therefore the corresponding value of  $r$  must be  $\frac{3}{10}$  of the way from .945 to .954. The difference between .945 and .954 is .009;  $\frac{3}{10}$  of .009 is about .003. Adding this to .945, we get .948 as a more precise value of the correlation coefficient by the second rank method. This is seen to differ slightly from .951, which we got by the footrule.

Notice that by both of the rank methods we get a value of  $r$  which is several points higher than .905, the value of  $r$  we got by the standard method. The two rank methods give results agreeing fairly closely with one another, but both are likely to differ appreciably from the true value of  $r$  by the standard method. (See the table on page 320.) For this reason the rank methods are not recommended except in case of non-linear relationship; i.e., those in which the line of relation is curved. In these cases the rank methods are presumed to show what the correlation would be if the line of relation were straight.

**EXERCISE 44.** Find by each of the two rank methods the correlation between other pairs of scores in Table 38.

**Correlation by unlike signs.** Sometimes it is not necessary to find a coefficient of correlation with accuracy. When only a rough approximation is needed, a coefficient may be found by the *method of unlike signs*.

This method takes account merely of the proportion of cases in which an individual stands above the median of the group with respect to one measure and below the median of the group with respect to the other. Thus if we call a score positive if it is above the median of its distribution, then in some pairs of scores both will be positive, in some both will be negative, and in some one will be positive and one negative. In the latter case the scores have like signs. (See Table 40.)

TABLE 40

	BELOW MED.	ABOVE MED.	
+	<div> <div>- +</div> <div>unlike signs</div> </div>	<div> <div>+ +</div> <div>like signs</div> </div>	ABOVE MED.
-	<div> <div>- -</div> <div>like signs</div> </div>	<div> <div>+ -</div> <div>unlike signs</div> </div>	BELOW MED.
	-	+	

The proportion of pairs of scores having unlike signs is a rough measure of the correlation. The higher the correlation, the lower this proportion tends to be. The relation between the proportion of unlike signs ( $U$ ) and the coefficient of correlation ( $r$ ) is shown in Table V in the Appendix.<sup>1</sup>

Let us suppose that we wished only a rough approximation of the coefficient of correlation in the case of the Ripon College

<sup>1</sup> When 50 per cent of the cases have unlike signs, the correlation is 00, of course; therefore the proportions of unlike signs extend only to .50. When more than 50 per cent of cases have unlike signs, the correlation is negative and the coefficient is found by taking the proportion of like signs.

Freshmen, Figure 55. In this case we should plot the pairs of scores as usual and then divide the two distributions as nearly at the median as possible. Thus, if we divide the Otis scores between the intervals 50-54 and 55-59, we have 53 above and 55 below the division; and if we divide the Terman scores between the intervals 160-169 and 170-179, we have 56 below and 52 above the division. These are near enough to equal divisions. (See Figure 56.)

There are 10 cases in the  $(-, +)$  section and 13 cases in the  $(+, -)$  section, making 26 cases of unlike signs. This number divided by 108, the whole number of cases, gives .21 as the proportion of unlike signs. Looking in Table V (page 297), we find the corresponding coefficient of correlation to be .79. The true coefficient you found to be .82, showing that the method is only approximate. The higher the correlation, the greater the probable error of a coefficient obtained by the unlike-sign method.

**EXERCISE 45.** Figure 57 is a portion of a curve showing the coefficient of correlation corresponding to each proportion of unlike signs from .00 to .20. Make a figure showing the curve extended to include the proportion of unlike signs from .00 to .50. (Plot points representing the pairs of values in Table V.)

**Other methods of measuring correlation.** There are several other methods of measuring relationship, most of which are used in the case of non-linear relationship — that is, when the line of relation between the two variables is known or believed to be curved.

One of these methods is the finding of the *correlation ratio* which is the ratio of the standard deviation of the means of the  $y$ -arrays to the standard deviation of the  $y$ 's themselves, or the ratio of the standard deviation of the means of the  $x$ -arrays to the standard deviation of the  $x$ 's. (See Kelley, *Statistical Method*, page 238.)



described briefly and is seldom used. (See Kelley, *Statistical Method*, page 265.)

The author has derived a formula for a coefficient of correlation, called a deviation formula, which yields the same value as that found by the Pearson product-moment method. The formula is

$$r = 1 - \frac{1}{2} \frac{\sigma_d^2}{\sigma_y^2}, \quad \text{(Formula 11, for a coefficient of correlation)}$$

in which  $d$  is the deviation of any point in the correlation plot above or below the line of relation (not the regression line). The calculation of a coefficient of correlation by a modification of this formula is illustrated in "The Reliability of Spelling Scales, Involving a 'Deviation Formula' for Correlation," *School and Society*, Vol. 4 (1916), Nos. 96 to 99, page 750. See also Arthur S. Otis and H. E. Knollin, "The Reliability of the Binet Scale and Pedagogical Scales," *Journal of Educational Research*, Vol. 4 (1921), page 132. This formula is useful with certain types of non-linear relationship. It serves as the basis for a convenient formula for a reliability coefficient of correlation (Formula 23, page 252).

#### QUESTIONS

1. Can you think of reasons why a coefficient of correlation calculated by means of the unlike-sign method would tend to differ somewhat from the coefficient calculated by the standard method?
2. Why do coefficients obtained by the rank methods also tend to vary somewhat from the coefficients obtained by the standard method?
3. Can you think of any traits that might be negatively correlated?

## CHAPTER SEVENTEEN

### CORRELATION AND PROGNOSIS

ONE of the chief purposes of correlation is for use as a measure of the degree to which one measure can be predicted from another. Thus, we may wish to know what success a student will probably have in studying a foreign language and may give him a "prognostic test." The amount of correlation that exists between scores in the test and subsequent scholarship in the study of the foreign language is a measure of the degree to which success or scholarship in the learning of the foreign language may be predicted by means of the test.

**The criterion.** In a case of this kind, some measure of scholarship must be used, of course, with which to correlate the scores. As a measure of relative success we may use the marks the students of a class have made in the final examination in the foreign language at the end of the year; or we may take the teacher's record of their daily recitations numerically expressed, or a combination of both. Whatever is used, this measure of success is called the *criterion*. The value of the test is judged principally by the degree to which it correlates with the criterion — the test in this case being given, of course, before instruction in the language was begun.

The correlation between prognostic tests in a foreign language and success in the study of the language, so far, has not been appreciably above .50. What does a coefficient of .50 signify in prognostic value of a test? Let us suppose that we have test scores ranging from 16 to 20 in our prognostic test and have the final-examination marks ranging from 60 to 100, and that the correlation between test scores and examination marks in the case of 64 students is as shown in Table 41. This is the same hypothetical correlation of .50 that we discussed at first. We are now merely viewing the meaning of correlation from one more angle.

TABLE 41

SHOWING THE HYPOTHETICAL CORRELATION BETWEEN SCORES IN A PROGNOSIS TEST IN FOREIGN LANGUAGE AND FINAL EXAMINATION MARKS

EXAMINATION MARKS	100			1	2	1
	90		2	6	6	2
	80	1	6	10	6	1
	70	2	6	6	2	
	60	1	2	1		
		16	17	18	19	20

SCORES IN PROGNOSIS TEST

Let us take first the four students who made the scores of 16. One made an examination mark of 60, two made 70, and one 80. What is the most probable examination mark a pupil would get who made a score of 16? We know from the line of relation drawn in Figure 48 that the mark that corresponds to the score of 16 is 60. But the tendency is always for students to get marks nearer to the mean of all the marks (in this case 80) than to the mark that corresponds to the score made. We can only say that the most probable mark a student would get who made a score of 16 is the mean of the marks received by those who made scores of 16. The mean mark of those who made 16 is 70. We should therefore *predict* that the student who made a score of 16 would make an examination mark of 70. But only two of the four made this mark. One made a score of 80. This is 10 points above the mean (70); so we should say that in this case the error of prediction was 10 points. Similarly, one of these four students made a score of 60; so we should say that in this case also the error of prediction was 10 points. The error of prediction in each case is equal to the distance

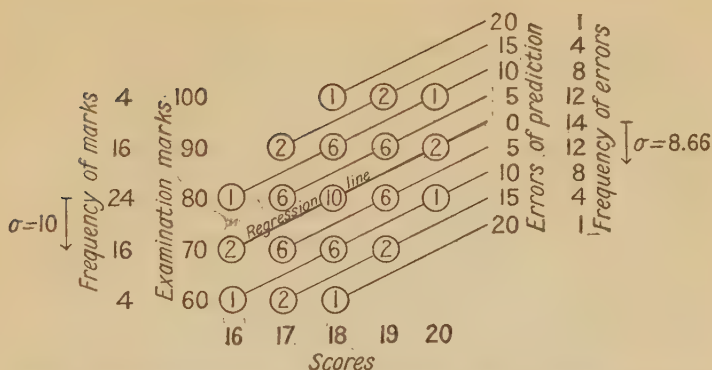


FIG. 58. Showing the finding of the distribution of errors of prediction in the case of a typical correlation plot.

of the point above or below the mean of the points in that array. (See Figure 58.)

Similarly, the most probable mark a student will get who has made a score of 17 is 75 — the mean of the marks of those who made 17. Therefore we should predict that a student who made a score of 17 would get an examination mark of 75. But no students actually got this mark. Six got 5 points more — the error of prediction in these six cases being, therefore, 5 points. Similarly, in six other cases the error of prediction was 5 points, and in four cases the error of prediction was 15 points. In each case the error of prediction is equal to the distance of the point above or below the regression line, since this passes through the means of the arrays. (See Figure 58.)

If we bring all these errors of prediction together, we find them distributed as shown at the right in Figure 58 and that the average error of prediction (disregarding direction of error) is  $6\frac{7}{8}$  points. You will see that if we made no use of the scores and merely predicted every mark to be 70 (the mean of all the marks), the average of our errors of prediction would

be no greater than (in fact, just equal to) the average deviation of the distribution of marks themselves. Now the average deviation of the distributions of examination marks themselves is only  $7\frac{1}{2}$  points, as you can readily see by working this out. Think what this means. It means that even though the correlation between scores and marks is .50, the average error of prediction of marks is still almost as great as the average of the errors one would make by mere chance guessing at the marks. For purposes of prognosis, therefore, a correlation of .50 is very low indeed.

We may think of the ratio of the average error of prediction of one measure from another to the average error of prediction that would result from pure chance guessing as a measure of the predictive value of the first measure. That is, the ratio of  $6\frac{7}{8}$  points to  $7\frac{1}{2}$  points is one measure of the predictive value of scores in our hypothetical case in predicting scholarship in foreign language. It is customary, however, to use for this purpose the ratio of the *standard* deviations instead of the ratio of the average deviation. The two ratios tend to be the same, however.

**The coefficient of alienation.** The standard deviation of the errors of prediction in Figure 58 is 8.66 points. (Notice that this is merely the standard deviation of the distances of the points above and below the regression line.) The standard deviation of the examination marks themselves is 10 points. The ratio of 8.66 to 10 is also a measure of the predictive value of the scores. This ratio is called the *coefficient of alienation*.<sup>1</sup> You will see that the greater this coefficient is, the less is the predictive value. We represent the coefficient of alienation by the letter *k*. If  $k = 1.00$ , there is no predictive value at all. If  $k$  is 0, prediction is perfect.

It is possible to calculate the value of *k* very simply from the coefficient of correlation for  $k = \sqrt{1 - r^2}$ . Thus we

<sup>1</sup> See T. L. Kelley, *Statistical Method*, page 173.

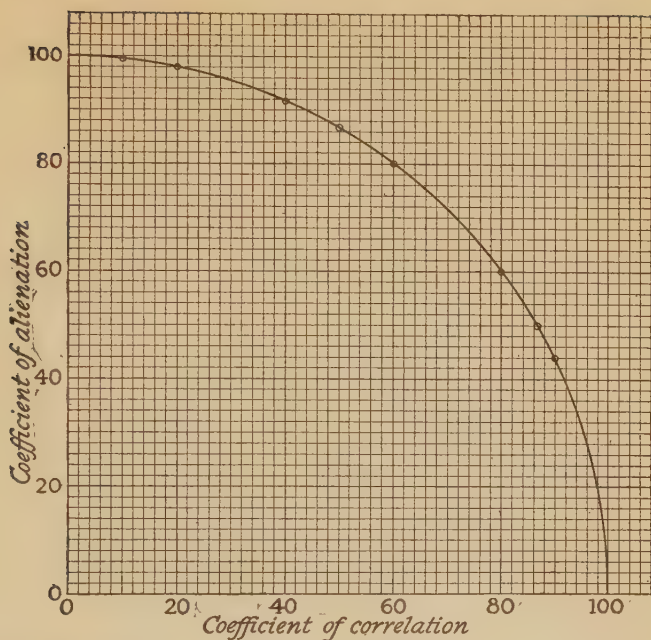


FIG. 59. Showing the relation between the coefficient of alienation and the coefficient of correlation.

might have told in advance that the ratio of the standard error (8.86) in Figure 58 to the standard deviation of examination marks (10) would be .886, for substituting the coefficient of correlation (.50) in the formula,  $k = \sqrt{1 - r^2}$ , we have  $k = \sqrt{1 - .50^2} = .886$ .

In order that you may see just what ratio the standard error of prediction of one variable from another bears to the standard error of prediction resulting from chance guessing in the case of correlations of various amounts, Figure 59 is given showing the relation between a coefficient of alienation and the corresponding coefficient of correlation. Note that the curve is in the form of a quarter of a circle.

TABLE 42

SHOWING THE CORRESPONDENCE BETWEEN COEFFICIENTS OF  
CORRELATION ( $r$ ) AND COEFFICIENTS OF ALIENATION ( $k$ )

$r$	$k$	$r$	$k$	$r$	$k$	$r$	$k$
1.00	.00	.85	.53	.70	.71	.51	.86
.99	.14	.84	.54	.69	.72	.49	.87
.98	.20	.83	.56	.68	.73	.47	.88
.97	.24	.82	.57	.67	.74	.46	.89
.96	.28	.81	.59	.66	.75	.44	.90
.95	.31	.80	.60	.65	.76	.41	.91
.94	.34	.79	.61	.64	.77	.39	.92
.93	.37	.78	.63	.63	.78	.37	.93
.92	.39	.77	.64	.61	.79	.34	.94
.91	.41	.76	.65	.60	.80	.31	.95
.90	.44	.75	.66	.59	.81	.28	.96
.89	.46	.74	.67	.57	.82	.24	.97
.88	.47	.73	.68	.56	.83	.20	.98
.87	.49	.72	.69	.54	.84	.14	.99
.86	.51	.71	.70	.53	.85	.00	1.00

*Problem:* Find the ratio of the standard error of prediction of pupils' examination marks from their scores in a prognosis test to the standard error that would be made by chance guessing, assuming that a correlation of .60 could be obtained between the test scores and examination marks. *Solution:* By consulting the curve in Figure 59 we see that a coefficient of correlation of .60 means a coefficient of alienation of .80. This means that the ratio called for would be .80.

EXERCISE 46. Find from Figure 59 the coefficient of alienation corresponding to coefficients of correlation of 20, 40, 50, 80, and 90. What coefficient of correlation would have to be obtained to reduce the standard error of prediction to  $\frac{1}{5}$  the standard error by chance guessing?

*Problem:* Make a table to show the coefficient of alienation corresponding to each coefficient of correlation from

1.00 down to .70. The values may be read from Figure 59. *Solution:* See Table 42. Note that the table works both ways; thus, a coefficient of alienation of 1.00 corresponds to a coefficient of correlation of .00, a coefficient of alienation of .99 corresponds to a coefficient of correlation of .14, etc. Therefore, it really was not necessary to extend the table beyond the value of .70 in each column.

**Further interpretation of a coefficient of correlation.** Table V gives us another aspect from which to interpret a coefficient of correlation. Let us refer again to our illustration of the prognosis test. Suppose that it is desired to accept for a class in a foreign language only one half of those who apply and those are selected who make scores in the prognosis test above the median. Suppose the correlation between prognosis test score and success in the course is .50. What proportion of those who would be excluded because of their test scores might have succeeded had they been allowed to take the course? By Table V we find the proportion of unlike signs corresponding to a coefficient of .50 to be  $.33\frac{1}{3}$ . This means that one third of the whole number of students would fall either in the group that would have been excluded but could have passed or in the group that would be admitted but would fail. And since these groups tend to be equal, it means that probably one third of those who had scores below the median in the prognosis test could pass the course and one third of those whose scores were above the median would fail it.

**EXERCISE 47.** Find from Table V what the correlation between prognosis test scores and scholarship should be so that the proportion of those whose scores were below the median but who could pass the course was only one fifth; one tenth.

**Interpretation of a coefficient of correlation in terms of displacement.** We are often interested to know how much displacement from a pupil's position in a group as the result of a

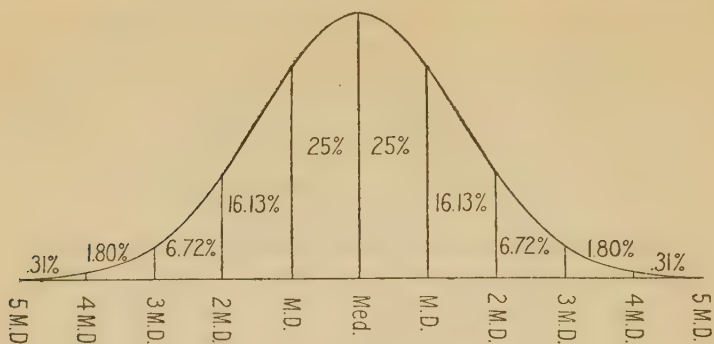


FIG. 60. Illustrating the division of the base of a normal surface of frequency into "tenths."

first test is likely to take place as the result of a second test, knowing the correlation between the two tests.<sup>1</sup>

To make the discussion concrete, refer to Figure 60, in which a normal surface of frequency is divided into 10 parts having equal widths, the width of each being the median deviation of the distribution. Note that 25 per cent of cases fall within either portion adjacent to the median of the distribution, 16.13 per cent fall within the next section on either side, and so on, down to the fifth section on either side, which contains .31 per cent of the distribution. These 10 sections, each 1 M.D. in width, therefore contain 99.92 per cent of cases, and for practical purposes we may consider these 10 M.D. distances as covering practically the whole range of a distribution. Or, in other words, we may consider each M.D. distance on the base line as one "tenth" of the distribution, and it is in this sense that we shall refer to "tenths" of the distribution in this illustration.

Now let us suppose that a pupil has made a score which will place him in the 6th "tenth." Let us suppose that he is given a second test that correlates with the first to the ex-

<sup>1</sup> This applies also to two forms of the same test.

TABLE 43

Showing, for various amounts of correlation, the chances in 100 that the second measure of an individual will be in the same "tenth"<sup>1</sup> of a distribution, not displaced more than 1 "tenth," 2 "tenths," etc. The sixth line is read as follows: In the case of a .50 correlation the chances are only 26 in 100 that an individual's second score will be in the same "tenth" of the distribution as his first score; the chances are 69 in 100 that his second score will not be displaced by more than 1 "tenth"; etc.

COEFFICIENT OF CORRELATION	NUMBER OF "TENTHS" DISPLACEMENT							
	0	1	2	3	4	5	6	7
.00	19	53	77	91	97	99.2	99.8	99.9 +
.10	20	55	79	92	98	99.4	99.9	
.20	21	58	82	94	98	99.6	99.9 +	
.30	22	61	85	95	99.0	99.8	99.9 +	
.40	24	64	88	97	99.5	99.9 +		
.50	26	69	91	98	99.7	99.9 +		
.60	29	73	94	99.2	99.9 +			
.70	34	81	97	99.8	99.9 +			
.80	41	89	99.2	99.9 +				
.90	55	98	99.9 +					
.95	71	99.9						
.98	91							
1.00	100							

tent of .50. What are the chances that his score in the second test will place him in the same "tenth" again; namely, the 6th "tenth"? Similarly, what are the chances that his score in the second test will place him either in the same "tenth" or in either of the two adjacent "tenths," and so on?

In order to answer these questions Table 43 is given. Note the line in the table opposite .50 in the column headed "Coefficient of Correlation." This line is read as follows: In the case of correlation of .50, there are 26 chances in 100 that the pupil mentioned above will make a score in the second test

<sup>1</sup> See explanation of special meaning of "tenth," page 224.

that will place him in the same "tenth"; there are 69 chances in 100 that his score will place him in either the same "tenth" or one or the other of the adjacent "tenths"; the chances are 91 in 100 that his score will place him in the same "tenth" or one or another of the two adjacent "tenths" on either side, and so on.<sup>1</sup>

Note that when the correlation is perfect there are 100 chances in 100 that there will be no displacement. This, of course, is obvious. On the other hand, when the correlation is .00, the displacement is as great as that which would take place by pure chance. In other words, there are 19 chances in 100 that the scores of any 2 pupils chosen at random would fall within the same "tenth"; there are 53 chances in 100 that the scores of these pupils would fall within two adjacent "tenths"; etc. This is what we should expect from zero correlation.

Remember that these "tenths" of a distribution are not deciles but are median deviation distances on the base line of the surface of frequency. Thus, if the median deviation of the distribution of scores in the first test is 5 points, then one "tenth" of the distribution as we are using the term here would mean an interval of 5 points in score.

**Errors of measurement.** All mental and educational tests are subject to what we may call *errors of measurement*. An error of measurement is the amount by which a particular score of an individual in a test deviates from his true score — the score he would have obtained if all conditions were exactly normal or standard. We may think of the true score as the median of a large number of scores.

<sup>1</sup> These values are general and refer to the chances of displacement, disregarding the position of the pupil's score. The chances for displacement vary somewhat for a single correlation, according to the position of the first score in the distribution. The farther the pupil's score from the median, the greater the chance of displacement. The values in Table 43 are average numbers of chances.

**The effect of errors of measurement upon correlation.** The coefficient of correlation between two variables is a measure of the degree to which a change in one tends to be accompanied by a change in the other. In all but the very exceptional cases of negative correlation this means, of course, that high values of one variable tend to be accompanied by high values of the other.

There are, of course, various factors which might operate to prevent two variables (such as ability in arithmetic and ability in reading) from being perfectly correlated. One is a difference between the inherent natures of the abilities, and another is the fact that measures of these abilities are not perfect. Thus, a pupil's score in the arithmetic test might happen to be in error downward and his score in the reading test in error upward, so that a difference in score due to inherent differences in the abilities measured might be increased by the errors. In other words, the errors of measurement tend to lower the coefficient of correlation between two variables. We call this lowering of a coefficient *attenuation*.

**Reliability.** If two forms of a mental-ability test measured exactly the same trait and measured it perfectly (consistently), the scores of a group of individuals in the two forms would show perfect correlation. The lack of perfect correlation between the two forms of a test is due principally to the errors of measurement. The greater the errors of measurement, the lower the correlation. The coefficient of correlation between two forms of a test, then, is a measure of the relative amounts of the errors of measurement of the test. In other words, the coefficient of correlation between two forms of a test is a measure of the *reliability* of the test.

The term *reliability* is used technically in connection with tests to mean the degree to which a test is consistent in measuring that which it measures. The various aspects of reliability are discussed by the author in an article entitled "The

Reliability of the Binet Scale and Pedagogical Scales.”<sup>1</sup> A coefficient of correlation between two forms of the same test or between two givings of the same test is called a *reliability coefficient*.

In exactly the same way that the coefficient of correlation between two forms of a test is reduced by errors of measurement, so is the coefficient of correlation between different kinds of tests reduced. Two tests cannot correlate with each other any higher than they correlate with themselves. That is, if each of two tests has a reliability coefficient of .80, these two tests cannot correlate with one another higher than .80 (heterogeneity being considered constant, of course). If they were correlated perfectly except for the errors of measurement, then they would be correlated to the extent of just .80 when subject to the errors of measurement.

**Correction for attenuation.** Let us designate two tests as  $X$  and  $Y$ . The reliability coefficients of the two tests may be called, respectively,  $r_{xx}$  and  $r_{yy}$  and the coefficient of correlation between them  $r_{xy}$ . In the case just cited in which  $r_{xx}$  and  $r_{yy}$  each equaled .80,  $r_{xy}$  cannot exceed .80. Let us suppose that  $r_{xx}$  equals .70 and  $r_{yy}$  equals .80 (heterogeneity of the group always assumed to be constant). The coefficient  $r_{xy}$ , then, cannot be higher than  $\sqrt{.70 \times .80}$ , no matter how nearly identical the two traits are that the tests measure. The coefficient  $r_{xy}$ , then, cannot exceed .748. If  $r_{xy}$  came out .748, we should know that except for the errors of measurement it would have been 1.00. If it came out only .70, we should know that the correlation between perfect measures of the two abilities would have been  $\frac{.70}{.748}$  of 1.00, or .935.

This value .935 is said to be corrected for attenuation due to errors of measurement — or, briefly, *corrected for attenuation*. To find how two tests would correlate theoretically if

<sup>1</sup> *Journal of Educational Research*, September, 1921.

each were a perfect measure of that which it measured, find the corrected coefficient by this formula :

$$r_{xy} \text{ (corrected for attenuation)} = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}. \quad (\text{Formula 12})$$

*Problem:* Suppose the reliability coefficients of two tests to be .75 and .84, and the correlation between them to be .61. What would this correlation be, corrected for attenuation?

*Solution:*

$$\frac{.61}{\sqrt{.75 \times .84}} = .77.$$

The correlation between the tests when corrected for attenuation is therefore .77.

See page 255 for further discussion of errors of measurement.

**EXERCISE 48.** If the reliability coefficients of two tests are .64 and .81 for a given group and the uncorrected coefficient of correlation between them is .70 for that group, what is the coefficient of correlation for that group between the abilities measured (the corrected coefficient)?

### QUESTIONS

1. In view of the foregoing discussion, what would you consider a satisfactory correlation between a prognostic test in stenographic ability or ability to learn a foreign language, and the criterion (ultimate success in the study in question)?

2. Can you think of any circumstances under which it might be desirable to find the coefficient between two variables corrected for attenuation?

3. Should a coefficient between a prognostic test and the criterion be corrected for attenuation?

## CHAPTER EIGHTEEN

### PARTIAL CORRELATION

It is not deemed to be within the scope of this textbook to give an extensive treatment of partial correlation, but an explanation of its meaning is desirable.

**A prevalent misconception regarding correlation.** We find a fairly high correlation between measures of brightness, such as the IQ, and ability to progress through school, and it is quite reasonable to assume that pupils who progress slowly are handicapped by dullness. When two variables are correlated, we tend to assume that one is the cause of the other.

We tend to believe, for example, that since there is a substantial correlation between pupils' progress through school and the general character of their homes, it must be that poor general home conditions must be an important factor in causing pupils to progress slowly through school. Now while it is undoubtedly true that certain home conditions, such as ungrammatical speech on the part of parents, do handicap a pupil in his progress through school to some extent, we might find progress through school correlated with other home conditions when there was no causal connection between them whatever.

Thus, if it is a fact (as seems probable) that true brightness is largely a matter of heredity and that differences in ability to progress through school are largely due to differences in brightness, and if it is a fact also (as seems probable) that "intelligent" parents tend to be those of higher economic status and to have better homes, then of course we should tend to find pupils who can progress rapidly through school coming from good homes, and pupils who progress slowly through school coming from poor homes. The result would be a positive correlation between pupils' progress through school and general home conditions. In other words, "in-

telligent " parents tend to have children who can progress rapidly through school and they tend to have " good homes."

We can easily imagine that homes of parents of higher economic status might tend to have a greater variety of dishes and cooking utensils, so that we would get a positive correlation between pupils' rates of progress through school and the variety of dishes and cooking utensils in the home. Of course one would not think of saying that paucity of dishes at home prevented pupils from progressing normally through school. This merely illustrates the fact, therefore, that correlation between " home conditions " and progress through school does not *prove* that one is in any sense the cause of the other.

All that is proved when two variables are found to be correlated is that some cause or causes (such as heredity) are operating to produce change in both variables. If we found a correlation of .60, for example, between scholarship in Latin and subsequent scholarship in English for a certain group of students, we could not infer any effect of the study of Latin upon the subsequent scholarship in English from that fact alone. It might be a case of the mentally mature students doing well in both Latin and English and the immature ones doing poorly in both.

**The need for a more direct measure of correlation.** What we should need in that case is a measure of the correlation between scholarship in Latin and scholarship in English when all the pupils are of the same degree of mental ability. Now it is not possible to get easily a large number of students of exactly the same mental ability, but there is a method of finding what the coefficient of correlation would be between scholarship in the two subjects if all the pupils were of the same mental ability. This is called the method of *partial correlation*.

Even with this coefficient, supposing it to be appreciable, we cannot say that the study of Latin affects subsequent scholarship in English. All we can say is that there is some

cause *other than differences in mental ability* operating to make those who stand well in Latin stand well subsequently in English. But even this is worth finding out.

Similarly, by the method of partial correlation we could find what the correlation would be between home conditions and ability to progress through school if all pupils were of the same degree of brightness. This might turn out to be a negligible correlation.

**How to find a coefficient of partial correlation.** The finding of a coefficient of partial correlation always involves three variables, such as scholarship in Latin (1), scholarship in English (2), and mental ability (3). It is customary to number these as shown, letting the two variables come first between which the coefficient of correlation is desired, with the effect of the third variable eliminated or held constant. Thus, if we wish to know what the coefficient of correlation is between scholarship in Latin and scholarship in English, with the effect of differences in mental ability eliminated (or made constant), we should let mental ability be numbered 3 and the others 1 and 2 in either order.

We then designate the coefficient of correlation between the first and second variables, with the effect of the third variable *not* eliminated, as  $r_{12}$ . We designate the coefficient of correlation between the first and third variables as  $r_{13}$  and that between the second and third variables as  $r_{23}$ . We designate the coefficient of correlation between the first two variables, with the effect of the third variable *eliminated*, as  $r_{12.3}$ .

The value of  $r_{12.3}$  is then found from the values of  $r_{12}$ ,  $r_{13}$ , and  $r_{23}$  by means of the following formula :

$$r_{12.3} = \frac{r_{12} - (r_{13} \times r_{23})}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}. \quad \begin{array}{l} \text{(Formula 13 for a coefficient} \\ \text{of partial correlation)} \end{array}$$

Let us suppose that the three variables mentioned in our illustration above were correlated as follows :

$r_{12}$ (Correlation between Latin and English)	= .60
$r_{13}$ (Correlation between Latin and Mental Ability)	= .70
$r_{23}$ (Correlation between English and Mental Ability)	= .80

To find the coefficient of correlation between scholarship in Latin and scholarship in English, with the effect of mental ability eliminated, we must solve the equation :

$$\begin{aligned}
 r_{12.3} &= \frac{.60 - (.70 \times .80)}{\sqrt{(1 - .70^2)(1 - .80^2)}} \\
 &= \frac{.60 - .56}{\sqrt{.51 \times .36}} \\
 &= \frac{.04}{.43} \\
 &= .09.
 \end{aligned}$$

According to our assumption, the correlation between scholarship in Latin and subsequent scholarship in English when the effect of differences in mental ability is removed is only .09 (a negligible correlation).

The partial correlation between two variables, the effect of a third having been eliminated, is sometimes said to be the correlation between the two variables, with the third variable "partialled out."

It is possible by means of other formulas to find the correlation between two variables with the effect of two others eliminated, or three others, or any number of others.

**Nature of partial correlation.** Let us assume for the sake of illustration that the correlation between scholarship in Latin and scholarship in English for pupils of the same degree of mental ability is zero, but that scholarship in each of these subjects is correlated with mental ability.

Let us assume that for 16 students whose mental ability is somewhat above the average for a certain class, the corre-

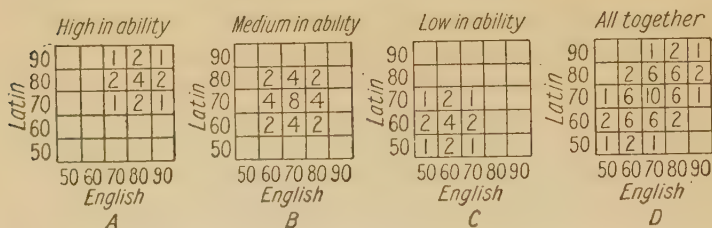


FIG. 61. Showing how two variables, of themselves uncorrelated, when subjected to the influence of a third variable show apparent correlation.

lation between scholarship in Latin and scholarship in English is as shown at *A* in Figure 61. This is, of course, a very artificial and over-simple scatter diagram. It represents exactly zero correlation.

Let us assume that for 32 students whose mental ability was just average for a certain class the correlation between scholarship in Latin and scholarship in English was as shown at *B* in Figure 61, and for 16 students below average for the class as shown at *C*. The scales are hypothetical. These correlations are also zero.

Now if we bring together all three of these groups of students into one group of 64 students for purposes of finding the correlation between scholarship in Latin and scholarship in English, the scatter diagram is that shown at *D*. Here we find the very same scatter diagram that was used in the first illustration of the finding of a coefficient of correlation, and, as shown there, the coefficient is .50.

Here, then, we have an apparent correlation of .50 between two variables known in advance to be totally uncorrelated!

We have already seen that if we did not know in advance how the two variables were correlated when not affected by the third but had to start with the correlation between the two variables as affected by the third, we could find the correlation between the first two variables without the influence

of the third by the method of partial correlation, provided we know the correlation of each with the third.

EXERCISE 49. Find the coefficients of partial correlation ( $r_{12.3}$ ) called for in the accompanying table.

$r_{12}$	.40	.50	.60	.75	.30	.40	.54	.83	.56	.41
$r_{13}$	.50	.50	.50	.50	.40	.40	.40	.58	.59	.13
$r_{23}$	.50	.50	.50	.50	.60	.60	.60	.23	.00	— .24
$r_{12.3}$										

**Effect of heterogeneity on correlation.** In exactly the same way as illustrated in Figure 61, the correlation between two measures such as scores in two mental-ability tests is affected by the amount of difference in age between the various pupils considered. Thus, let us suppose the correlation between ability in arithmetic and ability in reading for a given age (say 12 years) to be .50, as represented at *A* in Figure 62. Let us suppose the correlation between these abilities at a lesser age (say 11 years) to be also .50, but since the scores of this group will be lower (because of correlation between these abilities with age) we may represent the correlation plot by *B*. Let us suppose the correlation at 10 years to be also .50, but with still lower scores as shown at *C*. The scales of measurement again are hypothetical.

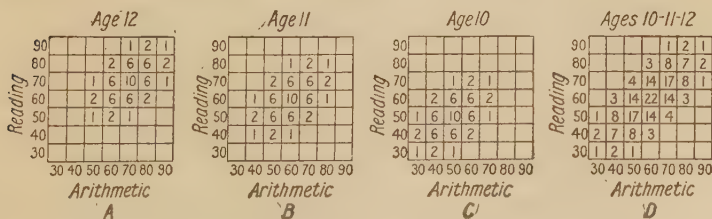


FIG. 62. Showing the effect of heterogeneity of the group upon a coefficient of correlation.

Now if we considered all three age groups together, we have a correlation plot as shown at *D*. If you should copy this plot to any position on the Otis Correlation Chart and work out the coefficient of correlation, you would find it to be .70. If you added Age 9 in the same way, the correlation would come out  $.77\frac{7}{8}$ , and if you added Age 8 also in the same way, it would come out  $.83\frac{1}{3}$ , etc., growing larger all the time.

Thus we see that when we are talking about the correlation between two abilities, each of which is affected by the age of the individuals, it makes a great difference whether the group is confined to a single age or contains individuals of various ages. The wider the range of ages, the higher the correlation will be. (See also page 254.)

*A statement of the correlation between two tests, therefore, without a statement such as the range of ages, or mental ages, or grades, to show the heterogeneity of the group, is valueless.* If the correlation between two tests is said to be .60 and no statement is made as to whether the pupils were all in one grade or ranged in grades from 3 to 9, you cannot tell whether, in case pupils had been of the same age or of the same mental age, the correlation would be .60, .40, .20, or even zero.

**How to correct a coefficient for a change in heterogeneity.** Suppose we know the coefficient of correlation between two tests to be .80 when the pupils were of various ages such that the standard deviation of scores in one test was, say, 15 points. Let us suppose that we find the standard deviation of scores in that test for a single age to be 10 points, and wish to know what the coefficient would be in the case of this more homogeneous group. The calculation is based on the ratio of these standard deviations<sup>1</sup> as follows:

<sup>1</sup>The ratio of any other measures of variability, such as the interquartile range, etc., will do, of course, since in all ordinary cases the ratio of any two similar measures of variability tends to be the same as the ratio of any other two similar measures of variability.

$$\begin{aligned}
 \text{New coef.} &= 1 - \left( \frac{\sigma_{\text{(old group)}}}{\sigma_{\text{(new group)}}} \right)^2 \times (1 - \text{old coef.}) && \text{(Formula 14)}^1 \\
 &= 1 - \left( \frac{15}{10} \right)^2 \times (1 - .80) \\
 &= 1 - 2.25 \times .20 \\
 &= 1 - .45 \\
 &= .55.
 \end{aligned}$$

This shows that the correlation between the two tests in the second case would be only .55.

We see from the above illustration that reducing the heterogeneity of a group so that the variability of scores is reduced to  $\frac{2}{3}$  will lower a coefficient of .70 to .55.

**EXERCISE 50.** If interested, you might work out the ratio of the standard deviations of either variable plots *A* and *D* in Figure 62 and see if you can show how the coefficient .70 might have been found by the above formula from the coefficient of .50, knowing this ratio.

Proof of the above formula is given in an article by the author, entitled, "Method of Inferring the Change in a Coefficient of Correlation Resulting from a Change in the Heterogeneity of the Group," in the *Journal of Educational Psychology*, May, 1922.

### QUESTION

If one author gives the reliability coefficient of his test as .80, and another author gives the reliability coefficient of his test as .90, would you say the second author's test was the more reliable?

<sup>1</sup> This formula is based on the assumption that the ratio of the standard deviations in one test is the same as the ratio of the standard deviations in the other test. The two ratios tend to be the same, of course, but if there were a slight difference in any particular case, it would cause the actual new coefficient to be slightly different from the estimated one calculated by the formula.

## CHAPTER NINETEEN

### MULTIPLE CORRELATION

**Measuring prognostic value.** One of the purposes to which tests are being put nowadays is the prediction of success in this or that undertaking. Thus, we give intelligence tests to predict general success in school work. We give specific prognosis tests to try to predict success in the study of foreign language, stenography, engineering, mechanical training, etc.

A measure of the value of a test in prognosis is the correlation between scores in the test and ultimate success in the study as measured by final examination marks, teachers' estimates of ability, final grades assigned by teachers, etc. Any one of these last-named measures that are used against which to judge the value of a test is called a *criterion*.

**The meaning of multiple correlation.** Let us suppose that we wish to predict success in learning stenography and have made up a test of quickness of forming certain associations which correlates to the extent of .60 with final examination marks which we have set up as our criterion. Not being satisfied with this, let us say, we have made up a test of a certain type of manual dexterity and found this to correlate only .50 with our criterion.

Do you think that a combined score in the two tests would correlate any higher with the criterion than either one alone? In other words, do you think the results of a test correlating .50 with a criterion would add anything of value to the results of a test that correlated .60 with the criterion? It is possible to determine this in advance. There is a formula for finding the correlation which it is possible to get between the combined score in two tests and a criterion, without actually combining the scores. This kind of correlation is called a *multiple correlation*.

In order to find the correlation between the combined score in two tests and a criterion, it is necessary first to find the correlation between the tests themselves. Let us suppose this to be .40. We may then substitute the three correlation coefficients in a formula and find the coefficient of multiple correlation.

**The formula for multiple correlation.** Let us call the association test Test 1, the dexterity test Test 2, and represent the criterion by  $C$ . The three coefficients of correlation will then be represented as follows :

Correlation between Criterion and Test 1 =  $r_{C1} = .60$

Correlation between Criterion and Test 2 =  $r_{C2} = .50$

Correlation between Test 1 and Test 2 =  $r_{12} = .40$

Let us represent the highest correlation that is possible between a combined score in Tests 1 and 2 and the criterion by the symbol  $R_{C.12}$ . Then

$$R_{C.12} = \sqrt{\frac{r_{C1}^2 + r_{C2}^2 - 2 r_{C1} r_{C2} r_{12}}{1 - r_{12}^2}} \quad \begin{array}{l} \text{(Formula 15}^1 \text{ for a} \\ \text{coefficient of mul-} \\ \text{tiple correlation} \\ \text{for three varia-} \\ \text{bles)} \end{array}$$

Substituting in this formula the values of  $r_{C1}$ ,  $r_{C2}$ , and  $r_{12}$  given above,

<sup>1</sup>Two formulas frequently used in finding a coefficient of multiple correlation are :

$$R_{C.12} = \sqrt{1 - (1 - r_{C2}^2)(1 - r_{C1.2}^2)} \quad \text{(Formula 16)}$$

$$\text{and} \quad R_{C.12} = \sqrt{1 - (1 - r_{C1}^2)(1 - r_{C2.1}^2)} \quad \text{(Formula 16 a)}$$

When both are used, the second serves as a check upon the first. It will be seen, however, that both of these formulas involve coefficients of partial correlation and for that reason Formula 15 is much shorter than either of these, since it involves only total correlations.

$$\begin{aligned}
 R_{C.12} &= \sqrt{\frac{.60^2 + .50^2 - 2 \times .60 \times .50 \times .40}{1 - .40^2}} \\
 &= \sqrt{\frac{.36 + .25 - .24}{.84}} \\
 &= \sqrt{\frac{.37}{.84}} \\
 &= .66.
 \end{aligned}$$

This shows that by properly combining the scores in Tests 1 and 2, which correlate .60 and .50 respectively with the criterion, and .40 with each other, it would be possible to get a correlation of .66 between the combined score and the criterion. The combined score, therefore, would be a slightly better prediction of success in learning stenography than either test alone.

**Finding the best weighting.** Remember that Formula 15 is a means of finding the *maximum* correlation with a criterion that can be obtained by combining two tests. If we merely added the raw scores in the two tests, we should not get the maximum correlation. Thus, in the case we are considering, to add the raw scores might yield a multiple correlation of only .63 or .64. In order to get the maximum correlation (.66 in this case), it is necessary that the scores in the two tests be made to contribute toward the combined score in proportion to their importance; that is, they must be *weighted*.

It is customary to use the raw scores in one test and to multiply the scores in the other test by some "weight" before combining them. This "weight" may be greater than 1 or less than 1, depending on whether the contribution of the second test needs to be increased or diminished. If we let  $S_T$  represent the total score of a pupil in the two tests when the second one is properly weighted so that the sum will give the best prediction of success, let  $S_1$  and  $S_2$  represent

the scores of the pupil in Tests 1 and 2 respectively, and let  $W$  represent the proper weight to give the second test, then <sup>1</sup>

$$S_T = S_1 + WS_2.$$

**Formula for weighting.** Now we can find the proper weight  $W$  from the three coefficients  $r_{C1}$ ,  $r_{C2}$ , and  $r_{12}$ , and the standard deviations  $\sigma_1$  and  $\sigma_2$  of the scores in the two tests by means of the following formula :

$$W = \frac{r_{C2} - r_{C1}r_{12}}{r_{C1} - r_{C2}r_{12}} \frac{\sigma_1}{\sigma_2}. \quad \begin{array}{l} \text{(Formula 17, for the} \\ \text{best weight)} \end{array}$$

Let us assume that  $\sigma_1 = 10$  and  $\sigma_2 = 12$ . Then, substituting the values of the coefficients and  $\sigma$ 's in Formula 15, we have

$$\begin{aligned} W &= \frac{.50 - .60 \times .40}{.60 - .50 \times .40} \times \frac{.10}{.12} \\ &= \frac{.26}{.40} \times \frac{.10}{.12} \\ &= .54. \end{aligned}$$

Therefore  $S_T = S_1 + .54 S_2$ .

This formula means that we should need to multiply the scores in Test 2 by .54 before adding them to the scores in Test 1 in order to get the best prediction of the criterion — best correlation of combined score with the criterion. The second test, therefore, deserves only a little more than half the weight of the first test. This is due partly to the fact that the correlation of the second test with the criterion is less than the correlation of the first test with the criterion, and partly to the fact that the standard deviation of scores in the second test was greater than the standard deviation of scores in the first test, a condition which of itself gave undue weight to the scores in the second test.

<sup>1</sup> This formula is read: "Total score = (First score) + (Weight  $\times$  Second score)."

**Symbolization.** We now know three kinds of correlation: total, partial, and multiple. Total correlation is ordinary correlation between two variables.

Total correlation between $x$ and $y$	$= r_{xy}$
Partial correlation between $x$ and $y$ , with $z$ constant	$= r_{xy.z}$
Multiple correlation between $x$ and $(y+ wz)$ [ $w$ = best weight]	$= r_{x.yz}$

Notice that in the case of a partial correlation a dot follows the first two subscripts (those of the two variables correlated) setting off the subscripts of the variable<sup>1</sup> that is made constant, while in the case of the multiple correlation the dot follows the first subscript (that referring to the criterion) setting off the subscripts<sup>2</sup> of the variables that are combined to correlate with the criterion.

**Relation of a multiple correlation to the total correlation.** If two tests correlate equally with a criterion, it is possible to make an estimate of the per cent by which the multiple correlation exceeds the correlation of the criterion with either test alone by means of Table 44. In that table "Gain in  $R_{C.12}$ " means the per cent that  $R_{C.12}$  exceeds  $r_{C1}$  or  $r_{C2}$  (these being equal).

Table 44 shows that when  $r_{12}$  is 1.00, the multiple correlation  $R_{C.12}$  is no greater than either  $r_{C1}$  or  $r_{C2}$ . This is another way of saying that nothing is to be gained by combining with one test another that correlates perfectly with it.

The table shows also that the less the value of  $r_{12}$ , the greater the gain of  $R_{C.12}$  over  $r_{C1}$  and  $r_{C2}$ . In other words, it is very important to combine tests that correlate with one another as little as possible, negatively if possible, to give the

<sup>1</sup> There may be two or more variables held constant. The correlation between variables  $a$  and  $b$  with the effect of variables  $c$ ,  $d$ , and  $e$  constant would be represented by  $r_{ab.cde}$ .

<sup>2</sup> There may be more than two variables combined. The correlation between  $a$  and the combination of variables  $b$ ,  $c$ ,  $d$ , and  $e$  would be represented by  $R_{a.bcd e}$ .

best multiple correlation (i.e., best combined correlation with the criterion). This is because the less the second test is like the first, the more it adds to the first when properly combined with it.

TABLE 44<sup>1</sup>

SHOWING THE PER CENT OF INCREASE OF  $R_{C,12}$  OVER  $r_{C1}$  OR  $r_{C2}$  FOR VARIOUS VALUES OF  $r_{12}$  IN THE SPECIAL CASE IN WHICH  $r_{C1} = r_{C2}$

$r_{12}$	1.00	.90	.80	.70	.60	.50	.40	.30	.20	.10
Gain in $R_{C,12}$	0.0%	2.6%	5%	8%	12%	15%	19%	24%	29%	35%
$r_{12}$	.00	-.10	-.20	-.30	-.40	-.50	-.60	-.70	-.80	-.90
Gain in $R_{C,12}$	41%	49%	58%	69%	83%	100%	124%	158%	216%	347%

**Regression equations.** In the foregoing discussion we have dealt with the finding of the best weight to give to a second measure before combining it with a first measure in order that the sum of the two thus weighted will correlate highest with a criterion, and with the finding of the correlation that would result from this best weighting between the combination of the two measures and the criterion.

This second purpose is the principal use to which multiple correlation is put. On rare occasions, however, it may be desirable to go farther and predict from the two measures of any individual an exact value of the criterion for that individual. This purpose calls for further calculation; namely, the solution of a *regression equation*.

For a criterion and two independent variables the regression equation is as follows:

$$\bar{X}_C = b_{C1.2} X_1 + b_{C2.1} X_2 + c. \quad \text{(Formula 18, the regression equation for three variables)}$$

<sup>1</sup>These relations hold, of course, only up to the point at which  $R_{C,12}$  becomes 1.00, which in general is reached before  $r_{12}$  becomes 1.00.

In this formula  $X_1$  stands for the first measure,  $X_2$  stands for the second measure, and  $\bar{X}_0$  stands for the value of the criterion predicted from the scores properly weighted. The values  $b_{C1.2}$  and  $b_{C2.1}$  are as follows:

$$b_{C1.2} = \frac{r_{C1} - r_{C2}r_{12}}{1 - r_{12}^2} \frac{\sigma_C}{\sigma_1}, \quad \text{(Formula 19, for a regression coefficient)}$$

$$b_{C2.1} = \frac{r_{C2} - r_{C1}r_{12}}{1 - r_{12}^2} \frac{\sigma_C}{\sigma_2}, \quad \text{(Formula 20, for a regression coefficient)}$$

$$\text{and } c = M_C - b_{C1.2}M_1 - b_{C2.1}M_2,$$

in which  $M_C$  = mean of values of variable C,

$M_1$  = mean of values of variable 1,

$M_2$  = mean of values of variable 2.

It is believed probable that the student will very seldom have occasion to build up a complete regression equation of this sort and that his purposes in practically all instances may be achieved by the use of the methods given above for weighting tests and for finding the coefficient of multiple correlation. For that reason the regression equation is given here merely for reference and no further explanation will be made of it. Those interested are referred to the more technical books mentioned on page 246.

Dr. Percival M. Symonds of Teachers College, Columbia University, New York, has drawn up a chart called the Symonds Multiple Correlation Chart to facilitate the calculation of coefficients of partial and multiple correlation and the coefficients necessary to build up the regression equation for any given case of three variables. The chart is analogous to the Otis Correlation Chart and consists of a "job analysis" of the steps involved in the necessary calculations, with spaces provided for making the calculations as in the Otis Correlation Chart. A reproduction of the chart is shown in Figure 63. Dr. Symonds has drawn up another chart for the case of four variables.

Symonds Multiple Correlation Chart  
(3 variables)

Variables		Coefficients		Means		St. Devs.	
0		$r_{01}$	1	$M_0$	4	$\sigma_0$	7
1		$r_{02}$	2	$M_1$	5	$\sigma_1$	8
2		$r_{12}$	3	$M_2$	6	$\sigma_2$	9

A		B		C		11 = $r_{01}r_{02}r_{12}$ 13 = $r_{02}r_{01}r_{12}$ 15 = $r_{12}r_{01}r_{02}$ 16 $1 - r_{01}^2$ 17 $1 - r_{02}^2$ 18 $1 - r_{12}^2$ 19 $B_{01.2}$ 20 $B_{02.1}$ 22 $r_{01.2}$ 24 $r_{02.1}$ 26 $r_{12.0}$ 27 $1 - r_{01.2}^2$ 28 $1 - r_{02.1}^2$ 31 $r_{0.12}$ 32 $\sigma_{0.12}$ 34 $b_{01.2}$ 36 $b_{02.1}$ 40 c
1		2		3		
2x3	10	1x3	12	1x2	14	
1-10	11	2-12	13	3-14	15	
D		E		F		
1 <sup>2</sup>	1.000	2 <sup>2</sup>	1.000	3 <sup>2</sup>	1.000	
1.000-1 <sup>2</sup>	16	1.000-2 <sup>2</sup>	17	1.000-3 <sup>2</sup>	18	
G		H		I		
11x18	19	13x18	20	17x18	21	
				$\sqrt{17x18}$	22	
J		K		L		
17x18	21	16x18	23	16x17	25	
$\sqrt{17x18}$	22	$\sqrt{16x18}$	24	$\sqrt{16x17}$	26	
M		N		O		
22 <sup>2</sup>	1.000	24 <sup>2</sup>	1.000	7x8	33	
1.000-22 <sup>2</sup>	27	1.000-24 <sup>2</sup>	28	19x33	34	
P		Q		R		
$\sqrt{29}$	32	7x9	35	7x9	35	
		20x35	36	20x35	36	
S		T		U		
34x5	37	36x6	38	37x8	39	
$\sqrt{34x5}$	40	37x8	39	4-39	40	

Regression equation  $X_0 = b_{01.2}X_1 + b_{02.1}X_2 + c$ 

or  $X_0 = (34)X_1 + (36)X_2 + (40)$

To find  $B_{01.2}$  use only Steps A, F, and G.To find  $B_{02.1}$  use only Steps B, F, and H.To find  $r_{01.2}$  use only Steps A, E, F, and I.To find  $r_{02.1}$  use only Steps N, D, F, and J.To find  $r_{12.0}$  use only Steps C, D, E, and K.To find  $r_{0.12}$  use only Steps A, E, F, G, I, L, and N.To find  $X_0$  use only Steps A, B, F, G, H, Q, R, S.

Designed and  
Copyright 1924 by  
Percival M. Symonds  
Teachers College  
Columbia University  
New York City, N.Y.

FIG. 63. The Symonds Multiple Correlation Chart.<sup>1</sup>

Dr. Truman L. Kelley has printed in his book, *Statistical Method*, a chart to facilitate the calculation of coefficients of partial and multiple correlation.

<sup>1</sup> Issued at present only in mimeograph form.

EXERCISE 51. Find the coefficients of multiple correlation ( $R_{C.12}$ ) called for in the accompanying table.

$r_{C_1}$	50	60	70	70	70	25	25	67	61	48
$r_{C_2}$	50	60	70	70	70	50	35	73	58	39
$r_{12}$	50	50	50	60	70	50	45	84	19	-27
$R_{C.12}$										

**Calculations with four or more variables.** In the foregoing discussion we have dealt only with the case of three variables. It is possible, however, to calculate coefficients of partial and multiple correlation and to build up regression equations using any number of variables. It is not deemed to be within the scope of this book, however, to consider these more complicated cases. The reader is referred to more technical books such as *An Introduction to the Theory of Statistics* by G. Udny Yule (C. Griffin & Co., London), and *Statistical Method* by Truman L. Kelley (The Macmillan Company, New York). Dr. Kelley also has a treatise on partial and multiple correlation in the *Handbook of Mathematical Statistics* by H. L. Rietz et al. (Houghton Mifflin Company, Boston), pages 139-149. Dr. B. R. Buckingham has written a very clear explanation of the meaning of partial correlation in an editorial in the *Journal of Educational Research* for April, 1923, and has explained multiple correlation in an editorial in the same journal for February, 1924.

The author has written an article entitled "The Derivation of Simpler Forms of Regression Equations," which appeared in the *Journal of Educational Psychology* for December, 1917. This article shows how to find the best weights when three and four tests are combined.

## CHAPTER TWENTY

### RELIABILITY

**The fallibility of test scores.** We can measure the height of a pupil and be reasonably sure that the measurement is correct, at least within a fraction of an inch. But suppose that we have given a 20-word spelling test to a group of pupils and Albert has made a score of 12. Is that the true measure of his spelling ability? Quite possibly not. It might be that if we had the true measure of his spelling ability in terms of scores in the test, we might find this true score to be as low as 9 or as high as 15. In other words, any single measure of a pupil's spelling ability may be in error (too high or too low) by 1, 2, 3, or possibly 4 points out of 20 points in score.

**Causes of variability of the scores of a test.** We may wonder why a pupil will not always make the same score in two equivalent forms <sup>1</sup> of a test. One reason for differences in scores of a given pupil is that any 20 words constitute only a sampling of the words the pupil may have studied, and the number of words that the pupil can spell that will happen to be among that 20 is subject to chance very much as the number of heads that will fall when 20 coins are flipped is subject to chance.

There are numerous other causes of variation of scores of a given individual in a given test, such as differences in attitude toward the test and effort put forth in taking it, etc. Sometimes, when the scoring calls for judgment on the part of the scorer, different scorers will score the same test paper in different ways.

**Reliability.** When a pupil makes scores very nearly the same, relatively speaking, in several equivalent alternative

<sup>1</sup> By equivalent forms is meant that each word in one form is matched by a word in the other form which would be correctly spelled by approximately the same percentage of a large number of pupils taking both tests.

forms of a test, the test is said to be *reliable*, but when a pupil makes scores in equivalent forms of the test that vary considerably, the test is said to be *unreliable*. Other things being equal, the less the variability of scores the greater the *reliability* of the test.

**Measures of reliability.** There are various ways of measuring the reliability of a test. One is to find the *probable error* of the test; another is to find the coefficient of correlation between two equivalent forms of the test. This coefficient is called the *reliability coefficient* of the test.

**Probable error of a test.** Suppose that our pupil, Albert, were given 25 equivalent alternative forms of a 20-word spelling test and made scores which, when distributed, were as follows:

Score . .	6	7	8	9	10	11	12	13	14
Frequency	1	2	4	3	5	4	2	3	1

Here the median score is 10 and we may assume for the moment that the score of 10 is the true measure of Albert's spelling ability in terms of scores in that test. If this were true, the three scores of 9 and the four scores of 11 are in error by 1 point and the other scores are in error as follows:

Error . .	0	1	2	3	4
Frequency	5	7	6	5	2

The median of these errors, as you see, is 2 points. This is the same as the median deviation of the distribution of 25 scores made by Albert. This median error might be used as a measure of the reliability of the test. That is, if some other test showed a median error of 3 points (other things being equal), we should consider the second test less reliable.

We should not judge the reliability of a test, however, on the basis of the scores of a single individual. Theoretically we should have a large number of scores of a large number of individuals. If we had enough scores of each individual so that we could consider the median score as the true score of that individual, then the deviations of the various scores of that individual from his median score would be the true amounts of error of his scores. And if we had the true values of the errors of many scores of many individuals in a test, the median of all these errors would be a very accurate measure of the reliability of the test. Such a median error is called the *probable error* of the test.

The *probable error* of a test, therefore, is the median value of the errors of many scores of each of many individuals; it is the most probable amount of the error of any new measurement in the sense that the error of the new measurement will as likely exceed that amount as not.

**How to find the probable error of a test.** Of course we cannot ordinarily give a test to a group of pupils a large number of times, but it is possible to obtain a theoretical value of the probable error of a test from the differences between pairs of scores of a sufficiently large group of pupils. This means that the test needs to be given but twice to each individual.

In fact, it is customary to give two equivalent alternative forms of the test to get two scores from each individual and to base the calculation of the probable error on these. Such a procedure assumes the true score of the pupil not as the median of a large number of scores in the same identical test but as the median of the scores of the individual in a large number of equivalent alternative forms of the test. This is a better assumption than the former, for if scores in two or more forms of a test are to be compared any deviation between them due to differences in the selection of

questions is in the nature of an error of measurement. Before finding the difference between the scores of any individual in the two forms of a test for this purpose, the effect of repetition of the test<sup>1</sup> and any difference in difficulty between them must be allowed for.

The formula for finding the probable error of a score from the differences between pairs of scores of single individuals. Let us suppose that we have found the distribution of differences between the two scores of each of the individuals of a group in a given test, all scores in the second form having been transmuted into terms of scores in the first form. *An approximate value of the probable error of the test may be found from the median value of these differences by simply multiplying by .707, which is  $\sqrt{\frac{1}{2}}$ .* That is,

$$\text{P.E.}_{(\text{score})} = \text{approximately } .707 (\text{Med. dif. bet. scores}).$$

(Formula 21, for the probable error of a score)<sup>2</sup>

A more accurate formula<sup>3</sup> is

$$\text{P.E.}_{(\text{score})} = .4769 \sigma (\text{dif. bet. scores}). \quad (\text{Formula 22})$$

**Need for other measures of reliability.** Suppose that we had given two forms of a 100-word spelling test to a group of pupils and had found the probable error to be 5 points. Should we say that the test was less reliable than the 20-word test? Obviously not, for an error of 5 points in 100 would probably mean considerably less displacement of a pupil in a rank order than an error of 2 points in 20. Thus, if the median deviation of the scores of a fifth grade in a 20-word spelling test was only 3 points, an error of 2 points might

<sup>1</sup> See "Practice effect," page 264.

<sup>2</sup> The proof of this formula and illustration of the method are given in an article by the writer entitled "The Reliability of the Binet Scale and of Pedagogical Scales" in the *Journal of Educational Research*, September, 1921, page 137. In this article is given also a convenient method of finding the median of the differences between scores in two forms of a test with differences in difficulty eliminated.

<sup>3</sup>  $.4769 = .6745\sqrt{\frac{1}{2}}$ .

change the percentile rank of a pupil from 50 to 33. Whereas if the median deviation of the scores of the fifth grade in the 100-word test was 15 points, an error of 5 points would probably not change a percentile rank of 50 more than 8 or 9 points. You can readily see that it makes a difference how variable the test scores are.

**Probable error in terms of variability of scores.** It would appear, then, that we should have a measure of the probable error in terms of the distribution of scores of the group under consideration. A simple way to do this would be to find what proportion the probable error is of the median deviation of the distribution of scores. If we did that in the cases we have just discussed, we should find that the probable error of scores in the 20-word spelling test was equal to  $\frac{2}{3}$  Med. Dev. (P.E. = 2, Med. Dev. = 3), whereas the probable error of scores in the 100-word test was equal to  $\frac{1}{3}$  Med. Dev. (P.E. = 5, Med. Dev. = 15). These values might not apply to a case in which some other group was tested, but for the group in question they are directly comparable and would show exactly how much more reliable a 100-word test was than a 20-word test, at least for the group tested.

**Reliability coefficient of correlation.** Another method of measuring the reliability of a test so as to take account of the variability of scores is to find the coefficient of correlation between two forms of the test; that is, between pairs of scores of a group of pupils in the test. This coefficient, as stated previously, is called the *reliability coefficient* of the test. It is sometimes called the coefficient of *self-correlation*. The higher the reliability coefficient, the greater the reliability of the test. A reliability coefficient of + 1.00 would mean perfect consistency. A reliability coefficient of .00 means that the two scores of any one pupil do not tend to be any closer together than the scores of two different pupils, chosen at random.

In the case of our hypothetical 20-word test with a P.E. of 2 points in score and median deviation of scores of 3 points, the self-correlation would be about .55, whereas the self-correlation in the case of the hypothetical 100-word test with a P.E. of 5 points and median deviation of score of 15 points would be about .89. Thus, while the number of points in the probable error of the 100-word test is greater, it is the more reliable of the two because relatively its probable error is less.

**Relation between reliability coefficient and probable error.** There is a very definite relation between the reliability coefficient and the probable error of a test when the latter is expressed in units of the median deviation of the distribution of scores. Thus, if we divide the probable error of a score by the median deviation <sup>1</sup> of the scores of the group under consideration, square this quotient, and subtract from 1.00, we have the reliability coefficient. Expressed in a formula,

$$\text{Rel. coef.} = 1 - \left( \frac{\text{P.E.}_{(\text{score})}}{\text{M.D.}_{(\text{dist'n})}} \right)^2. \quad (\text{Formula 23,}^2 \text{ for a reliability coefficient})$$

*Problem:* Knowing the probable error of a test to be 5 points and the median deviation of the distribution of scores in both forms of the test for a certain group of pupils to be 10 points, what would the reliability coefficient be in that case?

*Solution:* Rel. coef. =  $1 - (\frac{5}{10})^2 = .75$ .

The reliability coefficient would be .75.

**Formula for finding the probable error of a score from the reliability coefficient.** The rule also works backward, of course. If we know the reliability coefficient of a test for a given group and know the median deviation of the scores in the test (both forms), we can find the probable error of a score. The rule is to subtract the reliability coefficient from 1.00, take the square root of the difference, and multiply this

<sup>1</sup> Average of the median deviations of the two distributions.

<sup>2</sup> This formula is derived from Formula 11, page 216.

by the median deviation of the distribution of scores. Expressed in a formula,

$$\text{P.E.}_{(\text{score})} = \text{Med. Dev.}_{(\text{dist'n})} \times \sqrt{1.00 - \text{rel. coef.}}$$

(Formula 24, for the probable error of a score)

Remember that whenever the variability of a distribution is found by means of the percentile graph, the interquartile range divided by 2 may be used in place of the median deviation in all cases except where the distribution appears to be markedly skewed. That is, in all ordinary cases,

$$\text{Med. Dev.}_{(\text{dist'n})} = \left( \frac{75\text{-percentile} - 25\text{-percentile}}{2} \right).$$

*Problem:* Knowing the reliability coefficient of a test to be .91 for a given group and the interquartile range to be 20 points in each distribution of scores from which the reliability coefficient was derived, what is the probable error of the test?

$$\begin{aligned} \text{Solution:} \quad \text{P.E.} &= \frac{20}{2} \times \sqrt{1 - .91} \\ &= 10 \times \sqrt{.09} \\ &= 10 \times .3 \\ &= 3. \end{aligned}$$

The probable error of the test is 3 points in score.

**Interpretation of a reliability coefficient.** Just how reliable is a test that has a reliability coefficient of .75? The meaning of such a reliability coefficient may perhaps best be shown by interpreting it in the light of the relation between the probable error of a score and the median deviation<sup>1</sup> of the distribution of scores by which the reliability coefficient was found.

Table 45 will help to make this interpretation.

<sup>1</sup> Average of the two median deviations of the two distributions, which are presumably about equal.

TABLE 45

SHOWING THE CORRESPONDENCE BETWEEN THE RELIABILITY COEFFICIENT OF A TEST AND THE PROBABLE ERROR OF A SCORE IN TERMS OF (I.E., DIVIDED BY) THE MEDIAN DEVIATION OF THE DISTRIBUTION OF SCORES BY WHICH THE RELIABILITY COEFFICIENT WAS FOUND

Rel. Coef.	.00	.25	.50	.60	.70	.75	.80	.85	.90	.95	.98	.99	1.00
P.E. ÷ M.D.	1.00	.87	.71	.63	.55	.50	.45	.39	.32	.22	.14	.10	.00

By this table we can see that when the reliability coefficient of a test for a certain group of pupils is .75, the probable error of a score is just half (.50) as large as the median deviation of the scores themselves. This, then, is not a very high reliability coefficient.

**Reliability and heterogeneity.** The probable error of a score is entirely independent of the heterogeneity of any group by which it may be calculated. The value of the probable error of a score obtained from a homogeneous group will tend to come out exactly the same as that obtained from a heterogeneous group, since the difference between any two scores of a pupil tends to be the same whether he is in one group or another.

On the other hand, a reliability coefficient depends upon the heterogeneity of the group from which it is obtained. By studying the formula for the reliability coefficient on page 252, you will see that if we increase the heterogeneity of a group by including pupils of a wider range of ability, the median deviation of the distribution of scores becomes greater, the fraction  $\frac{\text{P.E.}}{\text{M.D.}}$  becomes less, and 1 minus the square of the fraction then becomes more. So the greater the heterogeneity of the group by which a reliability coefficient is calculated, the greater the latter will be. A reliability coefficient, therefore, without a statement as to the heterogeneity

of the group — number of grades included, or the like — is of little or no value as a measure of the reliability of a test.

In the same way it might be shown that if we could lessen the probable error of a test without changing the median deviation of scores for a given group, the reliability coefficient would be increased, and similarly an increase in the probable error would cause a decrease in the reliability coefficient.

**Errors of measurement.** We shall use the term *error of measurement* to refer to the amount by which a pupil's score in a test deviates from the hypothetical true score of that pupil — thinking of the true score as the score that would be obtained if all the conditions of the test were exactly normal or standard, or as the median of a large number of scores of that pupil in that test, the effect of repetition being eliminated. The probable error we have been considering is the theoretical median of these errors of measurement for any test. (See also page 226.)

It is easy to visualize the effect of errors of measurement on a reliability coefficient of correlation. Let us assume for the moment that the abilities measured by the two forms of a certain test are perfectly correlated. Let us suppose that a correlation plot showing the perfect correlation between true scores of 64 individuals in two forms of the test appeared as shown in Figure 64.

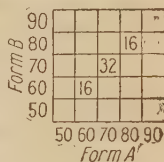


FIG. 64.

Let us now suppose that due to errors of measurement the 16 true scores of 80 in each test were scattered about as shown at *A* in Figure 61, page 231; that the 32 true scores of 70 in each test were scattered about as shown at *B* in Figure 61, and the 16 scores of 60 as shown at *C*.

The resulting correlation plot, of course, would appear as shown at *D* in Figure 61, which represents a correlation of .50.

Thus we see visually how errors of measurement, such as would scatter scores as shown at *A*, *B*, and *C*, in Figure 61,

will reduce a correlation of 1.00 as shown in Figure 64 to a correlation of .50.

**Validity.** There are two important aspects in which a test that purports to be a "standardized" test must be considered. One is with regard to reliability, and the other is with regard to what is called *validity*.

*The reliability of a test is the degree to which it is consistent in measuring that which it measures, but the validity of a test is the degree to which it measures that which it purports to measure.*

Thus, we may make up a test and call it a prognosis test in foreign language, with the idea of using it to predict whether students will succeed in the study of a foreign language. We have discussed tests of this type under the heading, "Correlation and prognosis," in Chapter XVII, page 217. The validity of the test would be the degree to which it did actually predict success in the study of a foreign language.

Now you can see that a test might be a very reliable measure of some ability, as for example the ability to spell foreign words phonetically, although this ability might not be so important in the study of foreign language that scores in the test would help much in predicting success. That is, the reliability coefficient of the "prognosis test" might be .80 while the correlation with actual subsequent success in the study of the language was only .50. The coefficient of .80 is a measure of reliability. The coefficient of .50 is a measure of validity.

**The probable error of a coefficient of correlation.** When we give two tests to the same group of pupils and find the coefficient of correlation between the two tests on the basis of the data, we have found what the degree of correspondence is between scores in the two tests for those pupils only. If we took another group of pupils as nearly like the first group as possible, or tested the same group again with the same

tests under conditions as nearly the same as possible, we should still get a slightly different coefficient. In fact, we might continue to take new groups of pupils or repeat the tests, and even though the groups or the conditions were as nearly alike as possible in every way, the coefficient would continue to vary a little, just as the scores of a single individual in a test might vary. Thus the coefficients of correlation between the National Intelligence Tests and Army intelligence examination, Alpha, might be, for 9 groups of 100 unselected adults, .80, .83, .79, .78, .81, .84, .76, .73, and .87. The median of these is .80, and the median deviation from the median is .03.

This value .03 is significant as to the variability of the coefficients, of course, since it shows that one half of the coefficients were within .03 of the median coefficient. If we had a very large number of coefficients of correlation between these two tests for groups of 100 unselected adults, we could safely assume the median of these to be most probably the true coefficient of correlation between these tests for unselected adults. It would most probably equal the coefficient of correlation that would be obtained if all the individuals were in one group.

The deviation of any coefficient from that median, therefore, may be considered as the amount by which that coefficient is in error. And the median of these deviations is, therefore, the median error of the coefficients.

Now, of course, we cannot find this median error because we do not have this large number of coefficients, but there is a formula by which the most probable value of that median error can be found, having given any single coefficient and the number of cases on which it is based. This most probable value of the median error is called the *probable error* of that coefficient *due to sampling*; that is, due to not having all possible cases.

The formula for the probable error of a coefficient. The formula for finding the probable error (P.E.<sub>r</sub>) of a coefficient (*r*) from the value of that coefficient is as follows :

$$\text{P.E.}_r = .6745 \frac{1 - r^2}{\sqrt{N}}. \quad \begin{array}{l} \text{(Formula 26, for the prob-} \\ \text{able error of a coefficient} \\ \text{of correlation)} \end{array}$$

To find the probable error of a coefficient, then, we subtract the square of the coefficient from 1.00, divide the remainder by the square root of the number of cases involved in finding the coefficient, and multiply the quotient by .6745.

You will see that the value *N* (number of individuals in the group) appears in the formula. An increase in the number of cases used to find a coefficient does not tend to increase or reduce the amount of the coefficient, so long as the individuals are chosen in the same way, but an increase in the number of cases does increase the *reliability* of the coefficient. That is, it reduces the probable error. The formula shows that the probable error is in inverse proportion to the square root of the number of cases ; that is, if we multiply the number of cases by 100 the P.E. will be  $\frac{1}{10}$  as great.

Solving the above formula for the P.E.<sub>r</sub> in the case in which *r* = .80 and *N* = 100 :

$$\begin{aligned} \text{P.E.}_r &= .6745 \frac{1 - .80^2}{\sqrt{100}} \\ &= .6745 \frac{.36}{10} \\ &= .024. \end{aligned}$$

While we found the median deviation of the distribution of nine hypothetical coefficients to be .03, the theoretical value of the median deviation of a large number of similarly obtained coefficients, calculated from the first one of the nine, is .024. In other words, the probable error of a coefficient of .80 based on 100 cases is .024. This is interpreted as meaning

that the chances are even that such a coefficient is correct within .024; that is, that it is between .776 and .824. We often say in a case such as this that the coefficient is  $.80 \pm .024$  (.80 plus or minus .024). If a coefficient is stated as being  $.50 \pm .05$ , it means that the coefficient is .50, with a probable error of .05.

How to find the probable error of a coefficient of correlation obtained by the Otis Correlation Chart. Provision is made on the Otis Correlation Chart for finding the probable error of a coefficient. This is merely the solution of the formula

$$\text{P.E.}_r = .67 \frac{1 - r^2}{\sqrt{N}}.$$

It consists of the following steps: (1) Square the coefficient, (2) subtract from 1, (3) take the square root of  $N$  (the number of cases), (4) divide the result in (2) by the result in (3), (5) multiply the quotient by .67 (or .6745 if desired).

EXERCISE 52. Find the probable errors of the coefficients of correlation found in Exercise 43. Use the spaces provided in the Otis Correlation Chart if this was used in Exercise 43.

**Significance of the probable error of a coefficient.** Suppose that we wish to know which of two prognosis tests is the better means of predicting what grades a student will make in a foreign-language course. Suppose the correlation between scores and grades in the case of one test is .50 and in the case of the other .55. May we say that the second is undoubtedly the better prognostic test?

That depends upon the probable errors of these coefficients. Suppose that the number of individuals in each case was 100. How would the probable errors of the coefficients compare with the difference between them?

The value of  $\text{P.E.}_r$  when  $r = .50$  and  $N = 100$  is .051, and when  $r = .55$  and  $N = 100$  it is .047. We see, therefore, that the difference between the coefficients is no greater than

the probable error of one of them. Hence it is not safe to say that the second test is more prognostic than the first. The experiment, at least, should be repeated. The average of two coefficients, each based on 100 cases, may be assumed to be as accurate as one coefficient based on 200 cases.

**The probable error of a difference.** We can go farther than merely to compare the difference between the two coefficients with their probable errors. We can find the probable error of the difference between the coefficients. This is exactly analogous to the probable error of a coefficient. It is the most probable value of the median deviation of the distribution of similar differences between the coefficient for one test and the coefficient for the other if we had a great many such pairs of coefficients.

The probable error of a difference<sup>1</sup> between uncorrelated measures depends on the probable errors of the two measures (in this case coefficients) as follows :

$$\text{P.E.}_{(\text{difference bet. two measures})} = \sqrt{\text{P.E.}_{(\text{one measure})}^2 + \text{P.E.}_{(\text{other measure})}^2}.$$

(Formula 27, for the probable error of the difference between uncorrelated measures)

Substituting the probable errors of our coefficients in this formula,

$$\text{P.E.}_{(\text{difference bet. two coeffs.})} = \sqrt{.051^2 + .047^2} = .069.$$

<sup>1</sup> **The standard deviation of a difference.** Formula 27 for the probable error of a difference between uncorrelated measures is derived from the more general formula

$$\sigma_{x-y} = \sqrt{\sigma_x^2 + \sigma_y^2 - 2 r_{xy} \sigma_x \sigma_y},$$

in which  $\sigma_x$  and  $\sigma_y$  are the standard deviations of the distributions of any two variables  $x$  and  $y$ , whether correlated or uncorrelated, and  $\sigma_{x-y}$  is the standard deviation of the distribution of  $x - y$ , the difference between them.

When the measures,  $x$  and  $y$ , are uncorrelated, then  $r_{xy} = 0$  and the formula reduces to  $\sigma_{x-y} = \sqrt{\sigma_x^2 + \sigma_y^2}$ , and Formula 27 follows directly from this formula.

Coefficients, in a case of this kind, are themselves uncorrelated, of course; so Formula 27 applies.

We see, therefore, that the difference between our two coefficients is less than the probable error of the difference between them. There is slightly more than a one-to-one chance, therefore, that the difference between the two coefficients is entirely accidental. Hence we should not attach much significance to this difference in the way of stating that one test is more prognostic than the other.

**Probable errors of a mean and of a standard deviation.** Suppose that we wish to find the norm for the sixth grade in the Stanford Achievement Test. Suppose we assumed the true norms to be the mean score of all sixth-grade pupils in the country. We cannot test all the sixth-grade pupils; so we test as many as we can and assume the group to represent the whole number of sixth-grade pupils. We assume the mean score of a group of 1000 sixth-graders to be fairly close to what the mean score of all sixth-graders would be, provided a representative selection was made. We assume also that the variability of the scores of our 1000 pupils will be fairly close to that of the whole number of sixth-graders. We have formulas, however, that tell us the "probable" amount by which a mean or standard deviation of a limited group will deviate from the same measure of the whole group, due to random sampling — that is, due to the fact that a portion of a group tends to deviate somewhat from the whole group in central tendency and variability.

Of course, if the selection of cases has not been made on a representative basis, there may be further deviation of the mean of the selected group from that of the whole group. There is no telling, for example, how far the mean score of a group of sixth-graders selected from New England only would differ from the mean of all sixth-graders. In this discussion only those deviations are considered that are due to "random sampling."

The probable error of the mean of a distribution due to random sampling is as follows :

$$\begin{aligned} \text{P.E.}_{(\text{mean of a dist.})} &= \frac{.6745 \times (\text{st. dev. of the dist.})}{\sqrt{\text{number of cases}}} \\ \text{or P.E.}_{(\text{mean})} &= \frac{.6745 \sigma}{\sqrt{N}}. \quad (\text{Formula 28, for the probable error of a mean}) \end{aligned}$$

The probable error of the standard deviation of a distribution due to random sampling is as follows :

$$\begin{aligned} \text{P.E.}_{(\text{st. dev. of a dist.})} &= \frac{.6745 \times (\text{st. dev. of the dist.})}{\sqrt{\text{number of cases}}} \\ \text{or P.E.}_{\sigma} &= \frac{.6745 \sigma}{\sqrt{N}}. \quad (\text{Formula 29, for the probable error of a standard deviation}) \end{aligned}$$

**Probable error of coefficients of partial and multiple correlation.** The probable error of a coefficient of partial or multiple correlation bears the same relation to the coefficient and the number of cases that the probable error of an ordinary coefficient does. That is,

$$\text{P.E.}(r_{12.3}) = .6745 \frac{1 - r_{12.3}^2}{\sqrt{N}} \quad (\text{Formula 30, for the probable error of a coefficient of partial correlation})$$

and

$$\text{P.E.}(r_{C.12}) = .6745 \frac{1 - r_{C.12}^2}{\sqrt{N}} \quad (\text{Formula 31, for the probable error of a coefficient of multiple correlation})$$

The probable error of a coefficient of correlation corrected for attenuation is so long and complicated that it is highly improbable that the reader would ever have occasion to use it. It is given in Kelley's *Statistical Method*, page 209.

**Reliability of averages.** If the reliability coefficient of a given test does not appear sufficiently high to justify reliance upon one score, it may be desirable to give two or more forms of the test to a single individual and to average the scores in

these forms. It is possible by means of a formula to determine in advance the reliability coefficient that would be obtained by finding the correlation between the average score in two forms of the test and the average score in two other forms of the test, or the reliability coefficient of average scores in any number of forms of the test. The formula is commonly known as "Brown's formula," but, as pointed out by Kelley,<sup>1</sup> this formula is only a special case of an earlier formula by Spearman for finding the correlation between the sum of the scores in one set of tests and the sum of the scores in any other set of tests.

If we let  $r_{aa}$  represent a correlation between the average score on  $a$  forms of a test and  $a$  other similar forms, and  $r_{1,1}$  represent the correlation between one form of the test and one other similar form, then

$$r_{aa} = \frac{ar_{1,1}}{1 + (a - 1) r_{1,1}}. \quad \text{(Formula 32, for the coefficient of correlation between the average score in } a \text{ forms of a test and } a \text{ other similar forms)}$$

In the special case in which we wish to find the correlation between the average score in two forms of a test and the average score in two other similar forms, we have merely to substitute 2 for  $a$  in the above formula. This gives us the formula

$$r_{2,2} = \frac{2 r_{1,1}}{1 + r_{1,1}} \quad \text{(Formula 33, for the correlation between the average score in two forms of the test and the average score in two other similar forms)}$$

*Problem:* Let us suppose that the reliability coefficient of a test when one form is correlated with one other form is .50. Find the reliability coefficient ( $r_{2,2}$ ) that would result from

<sup>1</sup> *Statistical Method*, page 205.

the correlation between the average score in two forms of the test and two other forms.

*Solution:* Substituting .50 for  $r_{1,1}$  in Formula 33,

$$\begin{aligned} r_{2,2} &= \frac{2 \times .50}{1 + .50} \\ &= \frac{1}{1.50} \\ &= .66\frac{2}{3} \end{aligned}$$

By this we see that by averaging two tests we may raise the reliability coefficient from .50 to  $.66\frac{2}{3}$ .

*Problem:* What would be the reliability coefficient ( $r_{3,3}$ ) if the average score in three forms was used?

*Solution:* In this case  $r_{1,1} = .50$  as before and  $a$  of Formula 32 = 3. Substituting these values in Formula 32,

$$\begin{aligned} r_{3,3} &= \frac{3 \times .50}{1 + 2 \times .50} \\ &= \frac{1.50}{2} \\ &= .75. \end{aligned}$$

This shows that if the average score in three forms of the test is used, the reliability coefficient is raised from .50 to .75.

**EXERCISE 53.** Find the reliability coefficient of average scores in two, three, and four forms of a test, the reliability coefficient of which when one form is used is .60; when it is .70; when it is .80.

**Practice effect.** One of the causes for differences between scores of an individual in a test, while not strictly classed as a source of unreliability, must nevertheless be taken into account. Thus we generally find that if a pupil is given an alternative form of a test within a short time after being given the first form, he will make an appreciably better score, even though the two forms are of the same degree of difficulty. Indeed, if a number of pupils take Form A of a test first and

Form B second, the majority will do better on Form B than on Form A ; whereas of another group who take Form B first and Form A second, the majority will do better on Form A than on Form B.

Obviously such a gain in score cannot be attributed to a difference in difficulty between forms. Presumably this gain or tendency to gain in a second form of a test is due principally to the pupils' being more familiar with the test the second time it is given. Thus, having gone over the test once, the pupils feel more familiar with it, understand the directions in advance and can therefore get under way more quickly ; and presumably they are less subject to distracting thoughts that may come to them when attacking a novel situation.

These various causes, whatever they may be, may be classed together and given the name *practice effect*. It is important to know what the practice effect is for any given test under given conditions.

To find the amount of practice effect for a given test after a given interval of time, the customary procedure is as follows: Give Form A first and Form B second to one group of pupils, and Form B first and Form A second to another group that corresponds very closely in ability to the first group. Find the median score of each group in each form of the test. Compare the medians of the two forms for the first group to see whether the median for the form given second is higher than the median for the form given first. Do the same for the two medians of the second group. If the forms of the test are of equal difficulty, the gain in the second median over the first will be the same for the two groups ; but if there is some difference in difficulty between the forms, the gain may be greater for one group than for the other and in that case find the average of the two gains. The effect of any difference in difficulty between the forms is in this way

eliminated, and the average number of points by which the median score in a second giving of a test exceeds the median score in the first giving may be considered as the practice effect of that test for that interval of time.

The practice effect of a test may be any number of points, from none at all up to 10 or 12 or possibly more, according to the test. The practice effect is much more marked in some types of tests than in others. The manual of directions accompanying a test should state the practice effect, in order that this may be taken into account when averaging the scores of a pupil in two forms of a test before comparing with norms. Thus, if a pupil has made a score of 20 in Form A of a test given first and a score of 24 in Form B of the test given second, and the practice effect of the test under the conditions under which it is taken is known to be 4 points, it would not be proper to find the average of the two scores 20 and 24 and compare this with the norms or with the first scores of other pupils, since norms are based on first scores, in which practice effect does not enter. If you wish to compare the average score of a pupil in the two forms of the test with norms or with first scores of other pupils, then 4 points (the practice effect) should be subtracted from the second score (24) before finding the average between the two scores, which are then both directly comparable with first scores. In other words, in order to compare the average of two scores of an individual with norms, the practice effect influencing the second score should first be eliminated by subtracting the number of points in practice effect from the second score before averaging the two scores.

#### QUESTION

What do you consider a satisfactory reliability coefficient?

## CHAPTER TWENTY-ONE

### GRADING AND CLASSIFYING

**The problem.** There seems to be no general understanding of just how mental and educational tests are to be used to accomplish what is perhaps the chief purpose for which they are designed — namely, so to grade and classify pupils for instructional purposes that they may be taught most effectively as well as most economically both as to their own time and effort and that of the teachers. This closing chapter will be devoted, therefore, to some of the general principles upon which grading and classifying are based.

Among the most important facts that we have learned from the use of mental-ability tests during the last decade or so are that pupils differ very much in ability to learn and that our schools have been so organized and conducted in the past that we find in the same grade in almost any school pupils who differ very much in their ability to learn. We have come to realize that this wide difference in ability to learn makes the work of the teacher very difficult and reduces the effectiveness of the instruction.

This is because the teacher in general can adjust the instruction only to the ability of those pupils who are about medium for the class and consequently there are two groups remaining for whom the instruction is more or less unsuited. There are those who find the work so easy that they could understand it with much less explanation than the teacher must give to the medium group and who are consequently not only wasting a great deal of time listening to this unnecessary explanation but are acquiring habits of laziness and of achieving less than the full amount of which they are capable.

There are also those who find the work so hard that they cannot understand it without more explanation and drill than

is necessary for the median group. The result is that the teacher must either take extra time to repeat explanations, thus wasting time of both the median and superior groups, or must allow these pupils to go without fully understanding the work. In that case they not only fall farther and farther behind the rest of the class but also become discouraged and come to think of themselves as failures.

This immediately raises the question as to why it is that pupils who find the work of their grade too easy are allowed to stay in it, and why other pupils get up into grades in which the work is too hard for them. Let us consider the first part of this question first. If the normal rate of progress of the normal pupil is one grade a year and a bright pupil can progress faster than this but is promoted only one grade a year, there will come a time sooner or later when the bright pupil will find himself in the fifth grade, for example, when he has already achieved as much as the normal sixth-grade pupil and can do successfully any sixth-grade work. The reason he is not placed in the sixth grade may be that these facts are not appreciated by the teacher, or it may be that it is felt that to place him in the sixth grade would mean that he will miss some important topic of instruction — he will not have “had Africa,” perhaps — or because it upsets the routine of the school somewhat to make such adjustments.

Now as to the slow pupils getting into grades too hard for them — this may be also because the teacher does not realize how much pupils of the same age differ in mental ability. She may think such a pupil is lazy, whereas he may be really incapable of doing the work of the grade.

It is the duty of every school principal first of all to know just what the condition of his school is in regard to the heterogeneity of ability within grades, then to learn whether any serious harm is done by shifting a pupil from one grade to another and to realize that the routine of the school cannot be

allowed to stand in the way of any pupil's getting the kind of instruction that is best suited to his ability.

Table 46 shows how widely the pupils of one school varied in achievement as measured by the Achievement Test. Notice that some pupils in Grade 4B did better than some in Grade 8A, and that the range of achievement in each grade is greater than the difference between the medians of the lowest and highest grades shown.

TABLE 46

SHOWING THE DISTRIBUTION OF ACHIEVEMENT (AVERAGE SCORES IN PART I OF FORMS A AND B OF THE OTIS CLASSIFICATION TEST) IN EACH OF VARIOUS GRADES IN A CERTAIN SCHOOL

SCORE	GRADE							
	4B	5A	5B	6A	6B	7A	7B	8B
90-94				1	1			
85-89				0	0	1	1	1
80-84				0	3	1	0	0
75-79				3	2	1	1	3
70-74				4	2	4	6	11
65-69			1	8	2	5	8	3
60-64			3	2	8	7	5	8
55-59	1	2	8	4	7	4	5	8
50-54	1	3	6	7	7	7	3	1
45-49	2	9	12	5	4	2	2	0
40-44	6	4	8	3	0	1	2	2
35-39	6	8	1	2	0			
30-34	8	5	0		0			
25-29	7	2	3		1			
20-24	3	1	1		1			
15-19	0	0						
10-14	1	1						

Table 47 shows even greater difference in mental ability with grades in the same school.

TABLE 47

SHOWING THE DISTRIBUTION OF MENTAL ABILITY (AVERAGE SCORES IN PART II OF FORMS A AND B OF THE OTIS CLASSIFICATION TEST) IN EACH OF VARIOUS GRADES IN A CERTAIN SCHOOL

SCORE	GRADE							
	4B	5A	5B	6A	6B	7A	7B	8A
70-74								1
65-69				4	2	2	3	2
60-64			1	0	5	4	4	7
55-59	1		4	9	7	5	8	7
50-54	0	2	5	6	5	7	6	6
45-49	2	4	5	5	6	6	4	2
40-44	1	5	8	4	2	1	4	4
35-39	7	4	3	5	4	4	0	1
30-34	2	4	6	1	4	3	1	
25-29	5	7	6	4	1	0	0	
20-24	6	4	2	1	0	1	1	
15-19	6	4	2	1	0			
10-14	5	0	1		1			
5-9		1			1			

Notice that one pupil in Grade 4B made an average score in the mental-ability test above the median for Grade 8B, and one pupil in Grade 7B made a score below the median for Grade 4B. Is it to be wondered that teachers find it hard to teach classes in a school such as this? Yet this is just the condition that exists in most schools until a reorganization is effected on the basis of standard tests.

**Regrading on the basis of mental ability.** Until recently conditions such as those shown in Tables 46 and 47 have been remedied by regrading chiefly on the basis of mental ability. Thus, when the regrading of a school was contemplated, it has been the custom to give a mental-ability test such as the National Intelligence Test. Then if a pupil in the fifth grade, for example, showed a mental ability above that nor-

mal for the sixth grade, he was considered a candidate for extra promotion into the sixth grade.

**Need for an achievement test.** When the actual shifting of pupils from one grade to another is to be done, the principal naturally seeks the judgment of the teacher as to whether the pupil's score is a true indication of his ability to do school work. More and more, teachers and principals are coming to realize that other data are needed for regrading than those furnished by a mental-ability test. There are factors other than mental ability determining the amount and character of the instruction a pupil may profit by and his fitness for this or that grade. Among these are probably interest, application, health, quality of previous instruction, etc. Now these factors may all affect the scores pupils will make in a test of actual achievement. For that reason an achievement test should be used in conjunction with a mental-ability test for purposes of regrading and classifying pupils.

It was to fill this need that the Otis Classification Test was devised. Provision is made for simply adding a pupil's score in the Mental Ability Test and the Achievement Test comprising the Otis Classification Test and calling the total score the Classification Score. If this were divided by 2, the result, of course, would be the average between the mental-ability score and the achievement score. And the distributions of scores in the two tests are sufficiently similar so that an average score can be found without transmuting scores of one into terms of the other. However, there is nothing to be gained by dividing by 2, since norms for the Classification Score are furnished. The total score is left as it is and fractions are avoided.

The distribution of Classification Scores of 525 pupils in Grades 3 to 9 in a Vermont school are shown in Table 48.

Here we see the usual overlapping of ability of pupils of the various grades that is found in practically every school that

TABLE 48

SHOWING THE DISTRIBUTIONS OF TOTAL SCORES IN THE OTIS CLASSIFICATION TEST (AVERAGE OF FORMS A AND B) IN VARIOUS GRADES OF ONE SCHOOL

SCORE	GRADE							TOTALS
	3	4	5	6	7	8	9	
160-169					1			1
150-159						1	2	3
140-149					2	7	15	24
130-139					3	8	13	24
120-129					8	10	17	35
110-119				3	12	16	10	41
100-109				4	14	5	7	30
90-99		1	3	8	9	13	2	36
80-89		1	2	9	11	5	1	29
70-79		2	13	14	3	3		35
60-69		3	15	18	7	2		45
50-59		5	13	13	2			33
40-49	2	12	20	7	0			41
30-39	6	28	6	1	1			42
20-29	15	15	7	2				39
10-19	33	7	1					41
0-9	24	2						26
Totals	80	76	80	79	73	70	67	525

has not given special attention to grading for homogeneous groups. Why should the pupil who made the score of 160 or more be in the seventh grade when his mental ability and actual achievement, according to total score in the Otis Classification Test, are greater than those of any pupil in the eighth or ninth grade? Similarly, if the pupil in the fourth grade who made a score of 90 or more is mentally as mature as any pupil in the fifth grade and can read, spell, and use language as well, knows as much geography, history, etc., and can solve problems in arithmetic as difficult as any pupil in the fifth grade, why should he be spending his time in the fourth grade?

Let us suppose that we wish to regrade the pupils of these seven grades in the middle of the school year so as to make them more nearly homogeneous within grades. There are, of course, both theoretical and practical considerations to be kept in mind. Let us consider the theoretical aspects of the problem first and then the practical ones.

**Regrading by counting papers.** If pupils could be promoted and demoted any number of grades at will, it would be a simple matter to make the pupils in each grade almost perfectly homogeneous in mental-educational ability (as we may call that which the total score in the Otis Classification Test measures). Thus one method would be to arrange the 525 papers in order of score, with the lowest on the bottom and the highest on the top. Then, if we wished to have the same number in each grade, we might divide 525 by 7 and finding that we could have 75 pupils in each grade, we might begin at the top of the pile and count off 75 papers and place those pupils in the ninth grade, count off another 75 papers and place those pupils in the eighth grade, etc. In that case our distributions of scores would appear somewhat as shown in Table 49.

**Regrading by a percentile curve.** After having found the distribution of scores of the 525 pupils by Table 48, there is a method of dividing them into homogeneous groups without arranging them all in one pile and counting out seven piles after which it might be necessary to sort them out again according to present grade.

Thus we might simply make a percentile curve to represent the whole distribution, as shown in Figure 65. By means of the scale 70 in Scale Chart A we could divide this curve horizontally into seven equal parts as shown and we should have immediately the range of scores for each new grade as shown in the figure.

TABLE 49

SHOWING THE THEORETICAL HOMOGENEOUS GROUPING OF PUPILS IN GRADES 3 TO 9 RESULTING FROM REGROUPING ON THE BASIS OF SCORE REGARDLESS OF THE AMOUNT OF SHIFTING OF PUPILS

SCORE	GRADES							TOTALS
	3	4	5	6	7	8	9	
160-169							1	1
150-159							3	3
140-149							24	24
130-139							24	24
120-129						12	23	35
110-119						41		41
100-109					8	22		30
90-99					36			36
80-89					29			29
70-79				33	2			35
60-69			3	42				45
50-59			33					33
40-49		2	39					41
30-39		42						42
20-29	8	31						39
10-19	41							41
0-9	26							26
Totals	75	75	75	75	75	75	75	525

Having found the range of scores for each new grade, we could go through the papers just as they are or take the names of the pupils just as they come in the Class Record and assign each immediately to the proper grade according to his score.

**A less drastic method of regrading.** The methods of regrading described above are, of course, drastic, not to say revolutionary. They illustrate the goal we might seek to reach in the way of homogeneous grading, but to achieve it all at once would necessitate both promotion and demotion of from one to three grades. It is preferable to work toward the goal gradually. Thus we might decide to promote at first

UNIVERSAL PERCENTILE GRAPH

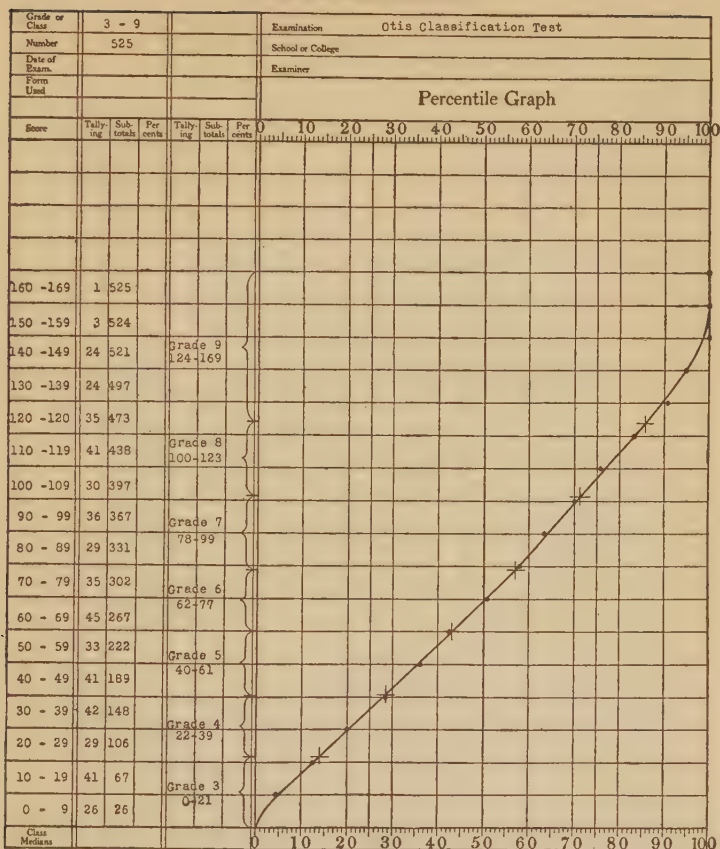


FIG. 65. Showing the method of regrading a school by means of a percentile curve.

only those whose scores were above the median, or even the upper-percentile score, of the grade above, and similarly to demote only those whose scores were below the median, or say the lower quartile, of the grade below.

### UNIVERSAL PERCENTILE GRAPH

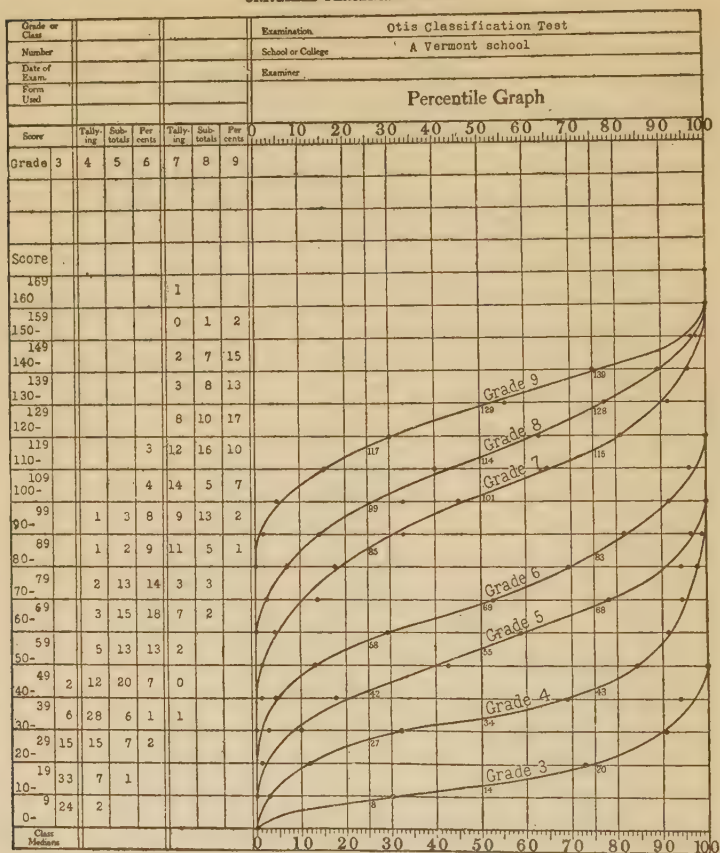


FIG. 66. Showing percentile curves representing the distributions of scores of Grades 3 to 9 of a Vermont school, in the Otis Classification Test.

In that case we should do well to draw percentile curves representing the distributions of scores of the seven grades separately, as shown in Figure 66. If we used medians, we could then promote from Grade 3 to Grade 4 all whose scores

exceeded 34, the median of Grade 4; from Grade 4 to Grade 5 all those in Grade 4 whose scores exceeded 55, etc. We should in that case demote from Grade 4 to Grade 3 any pupils in that grade whose scores fell below 14, the median of Grade 3, etc.

If we wished to be still more conservative, we might, as suggested above, promote from Grade 4 to Grade 5 only those whose scores exceeded 68, the upper-quartile score of Grade 5, and likewise demote from Grade 5 to Grade 4 only those whose scores fell below 27, the lower-quartile score of Grade 4, etc. Repetition of this procedure each year would effect a gradual change to homogeneous grouping so far as mental ability and achievement were concerned.

**Classification within grades.** In our discussion so far no account has been taken of the ages of the pupils. But if we should regrade our 525 pupils so as to have the most homogeneous groups from the standpoint of mental ability and achievement, as illustrated in Table 49, we should still find a wide range of ages within each grade, as shown in Table 50.

TABLE 50

SHOWING THE DISTRIBUTION OF AGES OF PUPILS HAVING SCORES WITHIN CERTAIN INTERVALS IN THE OTIS CLASSIFICATION TEST

GRADE	RANGE OF SCORE	AGES										TOTALS
		8 TO 8:11	9 TO 9:11	10 TO 10:11	11 TO 11:11	12 TO 12:11	13 TO 13:11	14 TO 14:11	15 TO 15:11	16 TO 16:11	17 TO 17:11	
9	124-169			1	3	13	29	17	10	2		75
8	102-123		1	0	10	16	25	13	8	2		75
7	78-101		1	10	12	13	14	12	7	3	3	75
6	62-77	1	5	14	21	19	6	6	2	1		75
5	40-61	6	6	20	16	17	6	3	1			75
4	22-39	60	20	18	13	10	5	1	2			75
3	0-21	24	19	16	7	5	3	0	1			75

Such a diversity of ages within single grades is generally considered ill-advised. The 15-year pupil, who belongs in Grade 3 because of his very immature mentality and extreme backwardness in knowledge of school subjects, feels, nevertheless, very much out of place in a class consisting chiefly of 8-, 9-, and 10-year-olds. Similarly, it is not considered suitable to place a 9-year child in the same class with pupils the majority of whom are adolescent, even though he may be their equal both in mental ability and knowledge.

For these reasons it is becoming the custom to classify pupils within the grade in such a way that the younger, brighter pupils are in one class and the older, duller pupils in another. If it is possible to make three classes within the grade, the younger, brighter third of the grade may constitute one class, the older, duller third another class, and the remaining, normal third another class. Or the bright and dull classes may each constitute about one quarter of the grade and the normal class the remaining half.

**Bright, normal, and dull sections.** You will note that we have used the expressions "younger, brighter pupils" and "older, duller pupils." The term "bright" in this connection is used to mean both mentally and educationally advanced for one's age. It is obvious that in a group of pupils having the same mental ability, the younger the pupil the brighter he is and the older the pupil the duller he is. When the pupils of a grade are of approximately the same degree of mental ability, as would be the case if graded as suggested in Tables 49 and 50, it would make very little difference, therefore, whether they were divided on the basis of age or on the basis of Classification Index (CI) (the measure used in connection with the Otis Classification Test, which is in the nature of an average between intelligence quotient and educational quotient).

Thus the 75 pupils in each grade might be divided into

classes so that the youngest 25 were in one class, the oldest 25 in another class, and the remaining 25 in another class. When making such a division, however, it is customary to place a little more than one third (say 30 in this case) in the bright section and a little less than one third (say 20 in this case) in the dull section, since teachers feel that dull pupils are harder to teach and that this arrangement equalizes the "teaching load." The result of such a classification in the case of our 525 pupils is as shown in Table 51.

TABLE 51

SHOWING THE METHOD OF CLASSIFYING PUPILS INTO BRIGHT, NORMAL, AND DULL SECTIONS (CONTAINING 30, 25, AND 20 PUPILS, RESPECTIVELY) ON THE BASIS OF AGE

GRADE	INTERVAL OF SCORE	CLASS	AGES										TOTALS
			8 TO 8:11	9 TO 9:11	10 TO 10:11	11 TO 11:11	12 TO 12:11	13 TO 13:11	14 TO 14:11	15 TO 15:11	16 TO 16:11	17 TO 17:11	
9	124-169	B N D			1	3	13	13 16	9 8	10	2		30 25 20
8	100-123	B N D		1		10	16	3 22	3 10	8	2		30 25 20
7	78-99	B N D		1	10	12	7 6	14	5 7	7	3	3	30 25 20
6	62-77	B N D	1	5	14	10 11	14 5	6	6	2	1		30 25 20
5	40-61	B N D	6	6	18 2	18	7 10	6	3	1			30 25 20
4	22-39	B N D	6	20	4 14	11 2	10	5	1	2			30 25 20
3	0-21	B N D	24	6 13	12 4	7	5	3	0	1			30 25 20

As suggested above, we might, if we wished, make the division on the basis of Classification Index, placing in the bright

section the third having the highest CI's, and in the dull section the third having the lowest CI's.

**The need for varied courses.** With the pupils graded and classified into bright, normal, and dull sections within each grade, as shown in Table 51, a school would be almost ideally organized so far as ease of teaching is concerned. The normal sections are the most homogeneous, of course, but the others are fairly so.

We must bear in mind, however, that the three sections of Grade 3 will not keep together throughout the remaining six years. In four years the normal section will presumably be in Grade 7. But the bright section, unless held back, should be at least in Grade 8 and possibly some of it in Grade 9. At any rate those pupils will have reached the mental ability and achievement of the normal pupils of these grades.

Similarly, the dull section will not have reached Grade 7 in four years unless the pupils have been promoted beyond their ability and achievement. The 15-year pupil in Grade 3, for example, may never get beyond Grade 5 even if he stays in school indefinitely, and the three pupils in Grade 3 who are 13 years old may not finish Grade 5 in less than four years.

For these reasons it has seemed desirable in many cases to allow the bright pupils to pursue a different course from that designed for the normal pupils, so planned that they may do the work of the eight grades in seven or six years or follow a more enriched curriculum, and similarly to allow the dull pupils to cover the work of only the first six or seven grades in eight years or to take only "minimum essentials."

**The three-track plan.** When such a need is felt, it is advisable to make the differentiation between bright, normal, and dull pupils on some basis before regrading and let these three groups follow separate courses. This scheme is called the *three-track plan*. These categories, of course, are not rigid and pupils may be moved at will from one to the other

whenever it seems desirable for any reason. In order to avoid the terms "bright," "normal," and "dull," it is better perhaps to call the three sections A, B, and C, or X, Y, and Z sections, as is done in Detroit.

Obviously this differentiation cannot be made on the basis of age, as it could be in the case of a group of pupils of about the same mental ability, since an 11-year pupil, for example, would be classed as "dull" if in Grade 3 and "bright" if in Grade 8. It must be made, therefore, on the basis of a measure of brightness such as the IQ. The Classification Index (CI) is better than the IQ, however, since it takes into account both mental and educational acceleration and retardation.

**The Interpretation Chart.** The most convenient method of finding a pupil's Classification Index is by means of the Interpretation Chart for the Otis Classification Test, as shown in Figure 67.<sup>1</sup> Thus, the uppermost dot in Figure 67 shows that a certain pupil was 11 years, 0 months old and made a score of 162 in the Otis Classification Test. The point lies almost on the curve marked 148, showing that the pupil had a Classification Index of approximately 148. (This, as has been explained, is the same degree of brightness that an IQ of 148 represents, but it takes into account actual achievement as well as mental ability.) To find a pupil's Classification Index, therefore, simply plot a point on the Interpretation Chart on the vertical line representing his age and on the horizontal line representing his score,<sup>2</sup> and note the number of the curve on which the point falls. This number is the CI of the pupil. If the point falls between curves, the CI may be estimated.

<sup>1</sup> Reproduced from page 28 of the Manual of Directions for the Otis Classification Test.

<sup>2</sup> Odd-numbered months are halfway between vertical lines, of course, and if the score is odd it is permissible to plot the point on the line representing the even score just below it, since these lines are so close together.

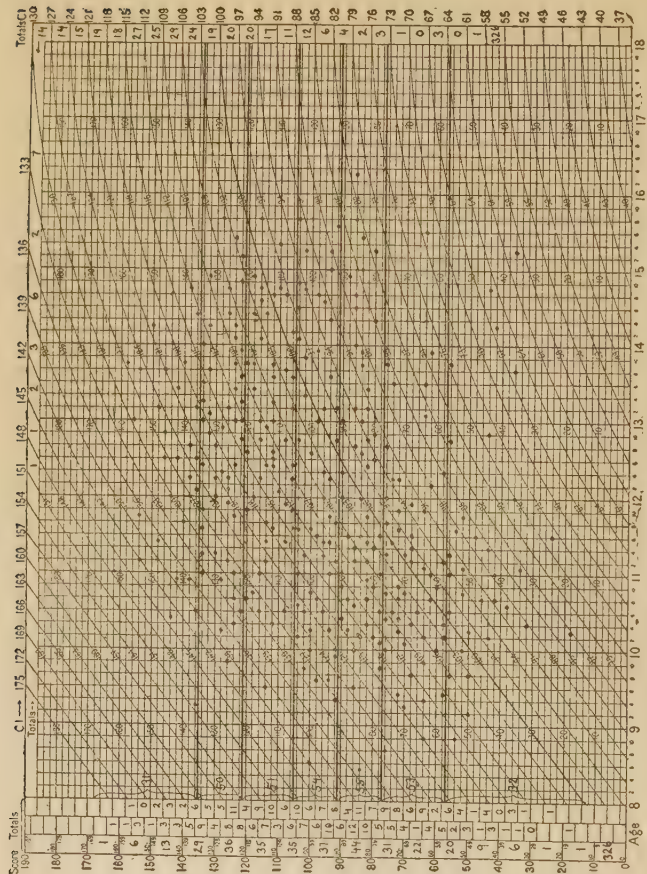


FIG. 67. A reproduction of the Interpretation Chart for the Otis Classification Test.  
(Actual size, 10 by 14 inches.)

**Convenience of the Interpretation Chart.** The value of the chart is that if a point is actually plotted when the CI of each pupil is being found, you have, when you have finished, a complete picture of the exact condition of your school, as shown in Figure 67, and can find the distribution of CI's in a very few minutes by simply counting the dots between each pair of curves. The distribution of CI's is shown at the top and the right side of the chart.

If this group of dots seems meaningless to you, you should study it carefully, for it is really very full of meaning. It shows you at a glance just how the pupils of any age vary in combined mental-educational ability and just how the pupils of any one level of mental-educational ability vary in age.

Thus the pupils between 11 and 12 years made scores varying from 24 to 162. And the pupils who made scores between 90 and 99 were of ages all the way from 9 years, 4 months to 15 years, 7 months. We have already given a table showing the distribution of ages for various intervals of score (Table 50), but, obviously, without the Interpretation Chart it would take a great deal of sorting, distributing, and counting to make this table and no CI's would be had then. As a by-product, so to speak, of the findings of the CI's, the general nature of these distributions is instantly apparent, however, from the Interpretation Chart.

**Classification for the three-track plan.** As explained above, we can find the distribution of CI's by counting the dots between each pair of adjacent curves. In so doing it is best to count those touching the lower curve but not those touching the upper curve. The CI's may then be considered to be grouped as follows: 100-102, 103-105, etc.

Having found the distribution of CI's as shown in Figure 67, we may then choose the intervals of CI to use in classifying the pupils into X, Y, and Z groups. Rather than to take the upper third of CI's for the X group and the lower third

for the Z group, it might be preferable to take the upper quarter and lower quarter for the X and Z groups, leaving one half the school in the Y group.

Let us suppose that we have chosen the quarter-half-quarter plan. One fourth of 525 is about 131. Counting down from the top, we find the upper 131 CI's to include 10 of the 32 in the interval 106-108; so we may decide that all pupils having CI's of 108 or over shall be placed in the X group, while those having CI's of 107 or 106 go into the Y group. Counting up from the bottom, we find the lower 131 CI's to include 6 of the 31 in the interval 85-87; so we may let all pupils having CI's below 86 be placed in the Z group. The remaining pupils, numbering about 263, will constitute the Y group.

**Grading within bright, normal, and dull groups.** The pupils constituting each of the three groups thus formed may then be divided into grades. The 263 or so pupils constituting the normal group will make seven grades of about 37 pupils each. This division may be made by arranging in a pile, in order of score, the papers of those pupils making CI's between 85 and 108 and counting them into seven piles. These pupils will constitute Grades 3 to 9 of the normal group, progressing on what we might call the main track.

The 130 or so pupils having CI's of 109 or over may be divided into any convenient number of grades on the basis of score and allowed to progress on what we might call the upper track. These pupils may either be given an enriched curriculum or be allowed to cover the work of a grade in less than a year.

Similarly the 130 or so pupils having CI's below 85 may be divided into any convenient number of grades on the basis of score and allowed to progress on what we might call the lower track. These pupils may be given only minimum essentials or allowed a year to cover less than a full year's work.

The advantage of the three-track plan is that a dull class may take just as much time as necessary to cover an ordinary year's work; or, to put it the other way, it may take the year to cover whatever portion of an ordinary year's work it can cover to best advantage, and it can go on from there next year regardless of what the normal group may be doing. And similarly the bright group may go as far in a year as it can to best advantage and go on from there the following year regardless of what the normal group is doing.

**No rigid classification.** As has been said, however, the division of pupils into so-called X, Y, and Z groups should never be considered final or permanent, and at any time that it seems desirable a pupil may be shifted from one group to another or back again. The purpose is merely to provide three different rates at which pupils may progress or instruction of three different intensities to fit better the widely varying abilities of the pupils and provide each with a course which he can pursue to best advantage — i.e., with least time lost, with maximum interest and effort, and with greatest benefit.

**Taking account of teachers' marks in grading and classifying.** In the above discussion of grading and classifying, no mention has been made of taking account of teachers' marks together with the pupils' scores in mental ability and achievement tests when regrading or classifying pupils. There is no reason, however, why teachers' marks should not be taken into consideration in such a case; and indeed it is very desirable to do so whenever possible, since the judgment of a competent teacher as to the ability of a child will always add something of value to the information that is obtained about him from the tests alone. There are always various personal traits that cannot be measured by tests, which, nevertheless, affect the child's ability to succeed in school.

**How to take account of teachers' marks.** The method of averaging scores and teachers' marks has been discussed

already in Chapter XI, and in order to make use of teachers' marks along with scores in mental ability and achievement tests it is necessary merely to combine these into a single measure which will correspond to the Classification Score.

To find the new combined score, convert the scholarship marks of each teacher into terms of Classification Score by the method explained in Chapter XI and find the average between the Classification Score of each pupil and his teacher's mark converted into terms of the Classification Score. As has been explained, any desired weight may be given to the teachers' marks. If it is desired to give the Classification Test double the weight given to teachers' marks, then in averaging the Classification Score and converted teachers' marks add the score twice and divide the sum by 3 in each case. The result may be treated for grading and classification purposes exactly as the Classification Score obtained from the test alone.

A measure, comparable to the Classification Index, which takes into account the teachers' marks may be obtained from the new combined score found as above by means of the Interpretation Chart in the usual manner. In other words, proceed with the new classification scores which take into account teachers' marks, in exactly the same way as with Classification Scores obtained from the test alone.

If combined measures of this sort are used in grading and classifying pupils, there is very little likelihood of any great error being committed, for the child will have been tested and rated from about every angle.

**Rating scale.** It is sometimes desirable to give numerical values to ratings of a subjective nature; that is, estimates made of qualities that cannot be measured by tests. For example, a teacher might wish to make estimates of the future scholarship of her pupils, based upon her judgment as to the effect that might result from some observed newly awakened

interest or ambition or change of environment or home conditions, or the like.

In a case of this kind it would not be possible, of course, for the teacher to obtain any precise measurement or even to make a precise estimate of such a future success, and it is therefore necessary to devise a scheme for converting subjective judgments into numerical terms.

It is customary to assume for this purpose that if the judgments of the teacher were expressed in true numerical terms, the measures of any fairly large group, such as the pupils of a class, would be distributed approximately according to the law of normal distribution. (See Figure 68.)

Now the middle 40 per cent of cases in a normal distribution cover a range of approximately 1 on the base line.<sup>1</sup> The 24 per cent on each side of the middle 40 per cent also cover a range of about 1  $\sigma$ , and the next 5½ per cent on either side cover a range of about 1  $\sigma$ . (See Figure 68.)

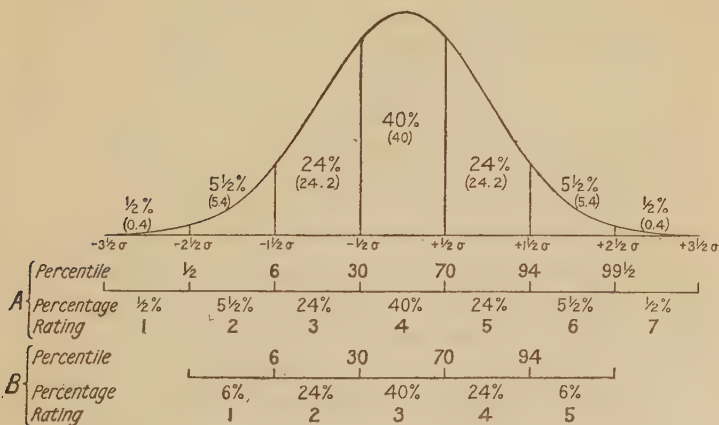


FIG. 68. Showing the method of making a rating scale.

<sup>1</sup> Actually from  $-.524\sigma$  to  $+.524\sigma$ .

Then there remains about  $\frac{1}{2}$  of 1 per cent on either side of these five intervals, practically all <sup>1</sup> of which lie within the next range of 1  $\sigma$ .

We may assume, therefore, that the seven groups marked off by the percentiles  $\frac{1}{2}$ , 6, 30, 70, 94, and  $99\frac{1}{2}$  are of approximately equal steps in true measure of the ability or trait estimated. We may give these groups the numerical values 1, 2, 3, 4, 5, 6, and 7, therefore, as shown at *A* in Figure 68. That is, any pupil who is estimated to be in the middle 40 per cent of the class in the trait estimated would be given a rating of 4; any pupil who is estimated to have a percentile rank in the class between 70 and 94 would be given a rating of 5; a pupil estimated to have a percentile rank between 94 and  $99\frac{1}{2}$  would be given a rating of 6; and a pupil estimated to be in the upper  $\frac{1}{2}$  of 1 per cent of the group would be given a rating of 7; and similarly for ratings 3, 2, and 1.

By the use of these assigned numerical values, the ratings may be combined with other measures by weighting them and averaging, or may be correlated with other measures, etc. No weighting is necessary for correlating purposes, but it might be desirable to use values 10, 20, 30, 40, 50, 60, and 70 for the seven ratings if these are to be averaged with scores, in order that they should be weighted more nearly in accordance with their value in comparison with the value of the scores with which they are combined.

**A "five-point" scale.** Unless there are about two hundred individuals in the group, it is not possible to use the seven-point scale just described, for  $\frac{1}{2}$  per cent represents only one individual in two hundred. For ordinary purposes, therefore, it is sufficient to use the five-point rating scale shown at *B* in Figure 68. Here the  $5\frac{1}{2}$ -per-cent and  $\frac{1}{2}$ -per-cent groups are merged into one 6-per-cent group, and ratings of 1, 2, 3, 4, and 5 only are given.

<sup>1</sup> Only about one hundredth of 1 per cent of a normal distribution lies above  $+3\frac{1}{2}\sigma$  or below  $-3\frac{1}{2}\sigma$ .

## APPENDIX ONE

### ABBREVIATIONS AND SYMBOLS

- Avg. Dev. or (A.D.) denotes average deviation. (See page 93.)
- A.R. or A.Q. denotes accomplishment ratio, previously called accomplishment quotient. (See page 173.)
- $b$  denotes a coefficient in a regression equation. (See page 243.)
- CI denotes Classification Index. (See page 281.)
- $d$  denotes a deviation. (See page 93.)
- G.S. denotes grade status. (See page 166.)
- $F$  denotes frequency of measures. (See page 196.)
- IQ denotes intelligence quotient. (See page 99.)
- I.Q.R. denotes interquartile range. (See pages 84 and 93.)
- $k$  denotes the coefficient of alienation. (See page 220.)
- $M$  denotes mean (sometimes median when meaning is clear from context as when used with  $Q_3$  and  $Q_1$ ). (See page 7.)
- Med. Dev. (M.D.) denotes median deviation. (See page 93.)
- Med. or Mdn. denotes median.
- $N$  denotes total number of cases. (See pages 7 and 186.)
- P.R. denotes percentile rank. (See page 26.)
- P.E. denotes probable error. (See pages 89 and 250.)
- P.E. with a subscript (i.e., P.E. <sub>$r$</sub> ) denotes the probable error of the quantity denoted by the subscript. (See page 258.)
- $\pm$  denotes "plus or minus." (See page 259.)
- $\pi$  (pi) denotes the ratio of the diameter of a circle to its radius.  
 $\pi = 3.14159$ . When used as an angle it denotes  $180^\circ$ . (See pages 211 and 290.)
- $Q$  denotes the value of the semi-interquartile range. [ $= (Q_3 - Q_1)/2$ .]  
(See page 34.)
- $Q_3$  denotes the upper quartile. (See page 34.)
- $Q_1$  denotes the lower quartile. (See page 34.)
- $r$  denotes a coefficient of correlation. (See page 189.)
- $r_{xy}$  denotes the coefficient of correlation between the two variables,  $x$  and  $y$ . (See page 242.)
- $r_{xx}$  denotes the coefficient of correlation between two forms of the same test. (See page 228.)
- $R$  denotes a measure of correlation found by the Spearman "Foot-rule," from which  $r$  may be found. (See page 210.)
- $r_{xy \cdot z}$  denotes a coefficient of partial correlation. (See page 242.)
- $r_{x \cdot yz}$  denotes a coefficient of multiple correlation. (See page 242.)

- $\rho$  (Rho) denotes a measure of correlation found by a rank method, from which the value of  $r$  may be found. (See page 211.)
- $\sigma$  (Sigma) denotes the standard deviation. (See page 93.)
- $\sigma_{1.2}$  denotes the standard deviation of the deviations of points in a scatter diagram from the line of regression. See *Statistical Method*, by Truman L. Kelley, page 155.
- $\Sigma$  (Sigma) denotes "sum of." (See pages 7 and 93.)
- S.R. denotes subject ratio. (See page 172.)
- $U$  denotes the proportion of unlike signs in a correlation plot. (See page 213.)
- $V$  denotes  $Y - X$ . (See page 197.)
- $x$ ,  $y$ , and  $z$  denote variables. (See page 193.)
- $X$ ,  $Y$ , and  $Z$  groups denote bright, normal, and dull groups. (See page 283.)

## THE GREEK ALPHABET

A $\alpha$	Alpha	I $\iota$	Iota	P $\rho$	Rho
B $\beta$	Beta	K $\kappa$	Kappa	$\Sigma$ $\sigma$	Sigma
$\Gamma$ $\gamma$	Gamma	$\Lambda$ $\lambda$	Lambda	T $\tau$	Tau
$\Delta$ $\delta$	Delta	M $\mu$	Mu	$\Upsilon$ $\upsilon$	Upsilon
E $\epsilon$	Epsilon	N $\nu$	Nu	$\Phi$ $\phi$	Phi
Z $\zeta$	Zeta	$\Xi$ $\xi$	Xi	X $\chi$	Chi
H $\eta$	Eta	O $\omicron$	Omicron	$\Psi$ $\psi$	Psi
$\Theta$ $\theta$	Theta	$\Pi$ $\pi$	Pi	$\Omega$ $\omega$	Omega

# APPENDIX TWO

## STATISTICAL TABLES

TABLE I  
PERCENTAGES OF NUMBERS FROM 11 TO 60

	16ths	17ths	18ths	19ths	20ths	21sts	22ds	23ds	24ths	25ths	26ths	27ths	28ths	29ths	30ths	
1	6	6	6	5	5	5	5	4	4	4	4	4	4	3	3	1
2	12	12	11	11	10	10	9	9	8	8	8	7	7	7	7	2
3	19	18	17	16	15	14	14	13	12	12	12	11	11	10	10	3
4	25	24	22	21	20	19	18	17	17	16	15	15	14	14	13	4
5	31	29	28	26	25	24	23	22	21	20	19	19	18	17	17	5
6	38	35	33	32	30	29	27	26	25	24	23	22	21	21	20	6
7	44	41	39	37	35	33	32	30	29	28	27	26	25	24	23	7
8	50	47	44	42	40	38	36	35	33	32	31	30	29	28	27	8
9	56	53	50	47	45	43	41	39	38	36	35	33	32	31	30	9
10	62	59	56	53	50	48	45	43	42	40	38	37	36	34	33	10
11	69	65	61	58	55	52	50	48	46	44	42	41	39	38	37	11
12	75	71	67	63	60	57	55	52	50	48	46	44	43	41	40	12
13	81	76	72	68	65	62	59	57	54	52	50	48	46	45	43	13
14	88	82	78	74	70	67	64	61	58	56	54	52	50	48	47	14
15	94	88	83	79	75	71	68	65	62	60	58	56	54	52	50	15
16	100	94	89	84	80	76	73	70	67	64	62	59	57	55	53	16
17		100	94	90	85	81	77	74	71	68	65	63	61	59	57	17
18			100	95	90	86	82	78	75	72	69	67	64	62	60	18
19				100	95	90	86	83	79	76	73	70	68	66	63	19
20					100	95	91	87	83	80	77	74	71	69	67	20
						100	95	91	88	84	81	78	75	72	70	21
							100	96	92	88	85	81	79	76	73	22
								100	96	92	88	85	82	79	77	23
									100	96	92	89	86	83	80	24
										100	96	93	89	86	83	25
											100	96	93	90	87	26
												100	96	93	90	27
													100	97	93	28
														100	97	29
															100	30
1	9	8	8	7	7											
2	18	17	15	14	13											
3	27	25	23	21	20											
4	36	33	31	29	27											
5	45	42	38	36	33											
6	55	50	46	43	40											
7	64	58	54	50	47											
8	73	67	62	57	53											
9	82	75	69	64	60											
10	91	83	77	71	67											
11	100	92	85	79	73											
12		100	92	86	80											
13			100	93	87											
14				100	93											
15					100											

Table I continued on pages 292-293.

	31sts	32ds	33ds	34ths	35ths	36ths	37ths	38ths	39ths	40ths	41sts	42ds	43ds	44ths	45ths	
1	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	1
2	6	6	6	6	6	6	5	5	5	5	5	5	5	5	4	2
3	10	9	9	9	9	8	8	8	8	8	7	7	7	7	7	3
4	13	12	12	12	11	11	11	11	10	10	10	10	9	9	9	4
5	16	16	15	15	14	14	14	13	13	12	12	12	12	11	11	5
6	19	19	18	18	17	17	16	16	15	15	15	14	14	14	13	6
7	23	22	21	21	20	19	19	18	18	18	17	17	16	16	16	7
8	26	25	24	24	23	22	22	21	21	20	20	19	19	18	18	8
9	29	28	27	26	26	25	24	24	23	22	22	21	21	20	20	9
10	32	31	30	29	29	28	27	26	26	25	24	24	23	23	22	10
11	35	34	33	32	31	31	30	29	28	28	27	26	25	25	24	11
12	39	38	36	35	34	33	32	32	31	30	29	29	28	27	27	12
13	42	41	39	38	37	36	35	34	33	32	32	31	30	30	29	13
14	45	44	42	41	40	39	38	37	36	35	34	33	33	32	31	14
15	48	47	45	44	43	42	41	39	38	38	37	36	35	34	33	15
16	52	50	48	47	46	44	43	42	41	40	39	38	37	36	36	16
17	55	53	52	50	49	47	46	45	44	42	41	40	40	39	38	17
18	58	56	55	53	51	50	49	47	46	45	44	43	42	41	40	18
19	61	59	58	56	54	53	51	50	49	48	46	45	44	43	42	19
20	65	62	61	59	57	56	54	53	51	50	49	48	47	45	44	20
21	68	66	64	62	60	58	57	55	54	52	51	50	49	48	47	21
22	71	69	67	65	63	61	59	58	56	55	54	52	51	50	49	22
23	74	72	70	68	66	64	62	61	59	58	56	55	53	52	51	23
24	77	75	73	71	69	67	65	63	62	60	59	57	56	55	53	24
25	81	78	76	74	71	69	68	66	64	62	61	60	58	57	56	25
26	84	81	79	76	74	72	70	68	67	65	63	62	60	59	58	26
27	87	84	82	79	77	75	73	71	69	68	66	64	63	61	6	27
28	90	88	85	82	80	78	76	74	72	70	68	67	65	64	62	28
29	94	91	88	85	83	81	78	76	74	72	71	69	67	66	64	29
30	97	94	91	88	86	83	81	79	77	75	73	71	70	68	67	30
31	100	97	94	91	89	86	84	82	79	78	76	74	72	70	69	31
32		100	97	94	91	89	87	84	82	80	78	76	74	73	71	32
33			100	97	94	92	89	87	85	82	80	79	77	75	73	33
34				100	97	94	92	89	87	85	83	81	79	77	76	34
35					100	97	95	92	90	88	85	83	81	80	78	35
36						100	97	95	92	90	88	86	84	82	80	36
37							100	97	95	92	90	88	86	84	82	37
38								100	97	95	93	90	88	86	85	38
39									100	98	95	93	91	89	87	39
40										100	98	95	93	91	89	40
41											100	98	95	93	91	41
42												100	98	95	93	42
43													100	98	96	43
44														100	98	44
45															100	45

	46ths	47ths	48ths	49ths	50ths	51sts	52ds	53ds	54ths	55ths	56ths	57ths	58ths	59ths	60ths	
1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1
2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	2
3	7	6	6	6	6	6	6	6	6	5	5	5	5	5	5	3
4	9	9	6	8	8	8	8	8	7	4	7	7	7	7	7	3
5	11	11	10	10	10	10	10	9	9	9	9	9	9	8	8	4
6	13	13	12	12	12	12	12	11	11	11	11	11	10	10	10	6
7	15	15	15	14	14	14	14	13	13	13	12	12	12	12	12	7
8	17	17	17	16	16	16	16	15	15	15	14	14	14	14	13	8
9	20	19	19	18	18	18	18	17	17	16	16	16	16	15	15	9
10	22	21	21	20	20	20	20	19	19	18	18	18	17	17	17	10
11	24	23	23	22	22	22	21	21	20	20	20	19	19	19	18	11
12	26	26	25	24	24	24	23	23	22	22	21	21	21	20	20	12
13	28	28	27	27	26	25	25	25	24	24	23	23	22	22	22	13
14	30	30	29	29	28	27	27	26	26	25	25	25	24	24	23	14
15	32	32	31	31	30	29	29	28	28	27	27	26	26	25	25	15
16	35	34	33	33	32	31	31	30	30	29	29	28	28	27	27	16
17	37	36	35	35	34	33	33	32	31	31	30	30	29	29	28	17
18	39	38	38	37	36	35	35	34	33	33	32	32	31	31	30	18
19	41	40	40	39	38	37	37	36	35	35	34	33	33	32	32	19
20	43	43	42	41	40	39	38	38	37	36	36	35	34	34	33	20
21	46	45	44	43	42	41	40	40	39	38	38	37	36	36	35	21
22	48	47	46	45	44	43	42	42	41	40	39	39	38	37	37	22
23	50	49	48	47	46	45	44	43	43	42	41	40	40	39	38	23
24	52	51	50	49	48	47	46	45	44	44	43	42	41	41	40	24
25	54	53	52	51	50	49	48	47	46	45	45	44	43	42	42	25
26	57	55	54	53	52	51	50	49	48	47	46	46	45	44	43	26
27	59	57	56	55	54	53	52	51	50	49	48	47	47	46	45	27
28	61	60	58	57	56	55	54	53	52	51	50	49	48	47	47	28
29	63	62	60	59	58	57	56	55	54	53	52	51	50	49	48	29
30	65	64	62	61	60	59	58	57	56	55	54	53	52	51	50	30
31	67	66	65	63	62	61	60	58	57	56	55	54	53	53	52	31
32	70	68	67	65	64	63	62	60	59	58	57	56	55	54	53	32
33	72	70	69	67	66	65	63	62	61	60	59	58	57	56	55	33
34	74	72	71	69	68	67	65	64	63	62	61	60	59	58	57	34
35	76	74	73	71	70	69	67	66	65	64	62	61	60	59	58	35
36	78	77	75	73	72	71	69	68	67	65	64	63	62	61	60	36
37	80	79	77	76	74	73	71	70	69	67	66	65	64	63	62	37
38	83	81	79	78	76	75	73	72	70	69	68	67	66	64	63	38
39	85	83	81	80	78	76	75	74	72	71	70	68	67	66	65	39
40	87	85	83	82	80	78	77	75	74	73	71	70	69	68	67	40
41	89	87	85	84	82	80	79	77	76	75	73	72	71	69	68	41
42	91	89	88	86	84	82	81	79	78	76	75	74	72	71	70	42
43	93	91	90	88	86	84	83	81	80	78	77	75	74	73	72	43
44	96	94	92	90	88	86	85	83	81	80	79	77	76	75	73	44
45	98	96	94	92	90	88	87	85	83	82	80	79	78	76	75	45
46	100	98	96	94	92	90	88	87	85	84	82	81	79	78	77	46
47		100	98	96	94	92	90	89	87	85	84	82	81	80	78	47
48			100	98	96	94	92	91	89	87	86	84	83	81	80	48
49				100	98	96	94	92	91	89	88	86	84	83	82	49
50					100	98	96	94	93	91	89	88	86	85	83	50
51						100	98	96	94	93	91	89	88	86	85	51
52							100	98	96	95	93	91	90	88	87	52
53								100	98	96	95	93	91	90	88	53
54									100	98	96	95	93	92	90	54
55										100	98	96	95	93	92	55
56											100	98	97	95	93	56
57												100	98	97	95	57
58													100	98	97	58
59														100	98	59
60															100	60

TABLE II

SHOWING THE PERCENTILE RANKS OF INDIVIDUALS MAKING VARIOUS SCORES MEASURED FROM THE MEDIAN IN TERMS OF THE MEDIAN DEVIATION (OR  $Q$ ) IN THE CASE OF A NORMAL DISTRIBUTION. A SCORE OF  $-5$  IS A SCORE 5 TIMES THE MEDIAN DEVIATION OR 5  $Q$  OF THE DISTRIBUTION *BELOW* THE MEDIAN

SCORE- $M$ $Q$	PERC. RANK	SCORE- $M$ $Q$	PERC. RANK	SCORE- $M$ $Q$	PERC. RANK	SCORE- $M$ $Q$	PERC. RANK
5.0	99.96	2.5	95.4	- 0.0	50.0	- 2.5	4.59
4.9	99.95	2.4	94.7	- 0.1	47.3	- 2.6	3.98
4.8	99.94	2.3	94.0	- 0.2	44.6	- 2.7	3.43
4.7	99.92	2.2	93.1	- 0.3	42.0	- 2.8	2.95
4.6	99.90	2.1	92.2	- 0.4	39.4	- 2.9	2.52
4.5	99.88	2.0	91.1	- 0.5	36.8	- 3.0	2.15
4.4	99.85	1.9	90.0	- 0.6	34.3	- 3.1	1.83
4.3	99.81	1.8	88.8	- 0.7	31.8	- 3.2	1.55
4.2	99.77	1.7	87.4	- 0.8	29.5	- 3.3	1.30
4.1	99.71	1.6	86.0	- 0.9	27.2	- 3.4	1.09
4.0	99.65	1.5	84.4	- 1.0	25.0	- 3.5	0.91
3.9	99.57	1.4	82.8	- 1.1	22.9	- 3.6	0.76
3.8	99.48	1.3	81.0	- 1.2	20.9	- 3.7	0.63
3.7	99.37	1.2	79.1	- 1.3	19.0	- 3.8	0.52
3.6	99.24	1.1	77.1	- 1.4	17.2	- 3.9	0.43
3.5	99.09	1.0	75.0	- 1.5	15.6	- 4.0	0.35
3.4	98.91	0.9	72.8	- 1.6	14.0	- 4.1	0.29
3.3	98.70	0.8	70.5	- 1.7	12.6	- 4.2	0.23
3.2	98.45	0.7	68.2	- 1.8	11.2	- 4.3	0.19
3.1	98.17	0.6	65.7	- 1.9	10.0	- 4.4	0.15
3.0	97.85	0.5	63.2	- 2.0	8.9	- 4.5	0.12
2.9	97.48	0.4	60.6	- 2.1	7.8	- 4.6	0.10
2.8	97.05	0.3	58.0	- 2.2	6.9	- 4.7	0.08
2.7	96.57	0.2	55.4	- 2.3	6.0	- 4.8	0.06
2.6	96.02	0.1	52.7	- 2.4	5.3	- 4.9	0.05

TABLE III

SHOWING THE VALUE OF  $r$  CORRESPONDING TO EACH VALUE OF  $R$  ACCORDING TO THE FORMULA  $r = 2 \cos (1 - R) 60^\circ - 1$  IN WHICH

$$R = 1 - \frac{6 \Sigma G}{n^2 - 1}$$

$R$	$r$	$R$	$r$	$R$	$r$	$R$	$r$
.01	.018	.26	.429	.51	.742	.76	.937
.02	.036	.27	.444	.52	.753	.77	.942
.03	.054	.28	.458	.53	.763	.78	.947
.04	.071	.29	.472	.54	.772	.79	.952
.05	.089	.30	.486	.55	.782	.80	.956
.06	.107	.31	.500	.56	.791	.81	.961
.07	.124	.32	.514	.57	.801	.82	.965-
.08	.141	.33	.528	.58	.810	.83	.968
.09	.158	.34	.541	.59	.818	.84	.972
.10	.176	.35	.554	.60	.827	.85	.975+
.11	.192	.36	.567	.61	.836	.86	.979
.12	.209	.37	.580	.62	.844	.87	.982
.13	.226	.38	.593	.63	.852	.88	.984
.14	.242	.39	.606	.64	.860	.89	.987
.15	.259	.40	.618	.65	.867	.90	.989
.16	.275-	.41	.630	.66	.875-	.91	.991
.17	.291	.42	.642	.67	.882	.92	.993
.18	.307	.43	.654	.68	.889	.93	.995-
.19	.323	.44	.666	.69	.896	.94	.996
.20	.338	.45	.677	.70	.902	.95	.997
.21	.354	.46	.689	.71	.908	.96	.998
.22	.369	.47	.700	.72	.915-	.97	.999
.23	.384	.48	.711	.73	.921	.98	.9996
.24	.399	.49	.721	.74	.926	.99	.9999
.25	.414	.50	.732	.75	.932	1.00	1.0000

TABLE IV

SHOWING THE VALUE OF  $r$  CORRESPONDING TO EACH VALUE OF  $\rho$  ACCORDING TO THE FORMULA  $r = 2 \sin (30^\circ \times \rho)$  IN WHICH  $\rho = 1 - \frac{6 \Sigma D^2}{n(n^2 - 1)}$

$\rho$	$r$	$\rho$	$r$	$\rho$	$r$	$\rho$	$r$
.01	.010	.26	.271	.51	.528	.76	.775+
.02	.021	.27	.282	.52	.538	.77	.785-
.03	.031	.28	.292	.53	.548	.78	.794
.04	.042	.29	.303	.54	.558	.79	.804
.05	.052	.30	.313	.55	.568	.80	.813
.06	.063	.31	.323	.56	.578	.81	.823
.07	.073	.32	.334	.57	.588	.82	.833
.08	.084	.33	.344	.58	.598	.83	.842
.09	.094	.34	.354	.59	.608	.84	.852
.10	.105-	.35	.364	.60	.618	.85	.861
.11	.115+	.36	.375-	.61	.628	.86	.870
.12	.126	.37	.385+	.62	.638	.87	.880
.13	.136	.38	.395+	.63	.648	.88	.889
.14	.146	.39	.406	.64	.658	.89	.899
.15	.157	.40	.416	.65	.668	.90	.908
.16	.167	.41	.426	.66	.677	.91	.917
.17	.178	.42	.436	.67	.687	.92	.927
.18	.188	.43	.446	.68	.697	.93	.936
.19	.199	.44	.457	.69	.707	.94	.945+
.20	.209	.45	.467	.70	.717	.95	.954
.21	.219	.46	.477	.71	.726	.96	.964
.22	.230	.47	.487	.72	.736	.97	.973
.23	.240	.48	.497	.73	.746	.98	.982
.24	.251	.49	.507	.74	.756	.99	.991
.25	.261	.50	.518	.75	.765+	1.00	1.000

TABLE V

SHOWING THE COEFFICIENT OF CORRELATION ( $r$ ) CORRESPONDING TO EACH OF VARIOUS PROPORTIONS ( $U$ ) OF UNLIKE SIGNS

$U$	$r$	$U$	$r$	$U$	$r$	$U$	$r$
.00	1.0000	.13	.917	.26	.685	.38	.368
.01	.9996	.14	.904	.27	.662	.39	.339
.02	.9982	.15	.891	.28	.638	.40	.309
.03	.9958	.16	.876	.29	.613	.41	.279
.04	.9924	.17	.860	.30	.588	.42	.249
.05	.998	.18	.844	.31	.562	.43	.218
.06	.983	.19	.827	.32	.536	.44	.188
.07	.976	.20	.809	.33	.509	.45	.156
.08	.969	.21	.790	.33 $\frac{1}{3}$	.500	.46	.125
.09	.960	.22	.771	.34	.482	.47	.094
.10	.951	.23	.750	.35	.454	.48	.063
.11	.941	.24	.729	.36	.426	.49	.031
.12	.930	.25	.707	.37	.397	.50	.000

TABLE VI

PROBABLE ERROR OF A COEFFICIENT OF CORRELATION ( $r$ ) DUE TO SAMPLING FOR VARIOUS VALUES OF  $r$  AND FOR VARIOUS NUMBERS OF MEASURES ( $N$ )

N (NUM. MEAS.)	VALUES OF $r$ (COEFFICIENT OF CORRELATION)												
	.00	.10	.20	.30	.40	.50	.60	.70	.75	.80	.85	.90	.95
10	.213	.211	.205	.194	.179	.160	.137	.109	.093	.077	.059	.041	.0208
20	.151	.149	.145	.137	.127	.113	.097	.077	.066	.054	.042	.029	.0147
30	.123	.122	.118	.112	.103	.092	.079	.063	.054	.044	.034	.023	.0120
40	.107	.106	.102	.097	.090	.080	.068	.054	.047	.038	.030	.020	.0104
50	.095	.094	.092	.087	.080	.072	.061	.049	.042	.034	.026	.018	.0093
60	.087	.086	.083	.079	.073	.065	.058	.044	.038	.031	.024	.017	.0085
70	.081	.080	.077	.073	.068	.060	.052	.041	.035	.029	.022	.015	.0079
80	.075	.075	.072	.069	.063	.057	.048	.038	.033	.027	.021	.014	.0074
90	.071	.070	.068	.064	.060	.053	.045	.036	.031	.026	.020	.014	.0069
100	.067	.067	.065	.061	.057	.051	.043	.034	.029	.024	.019	.013	.0066
125	.060	.060	.058	.055	.051	.045	.039	.031	.026	.022	.017	.011	.0059
150	.055	.055	.053	.050	.046	.041	.035	.028	.024	.020	.015	.010	.0054
200	.048	.047	.046	.043	.040	.036	.031	.024	.021	.017	.013	.0091	.0047
250	.043	.042	.041	.039	.036	.032	.027	.022	.019	.015	.012	.0081	.0042
300	.039	.039	.037	.035	.033	.029	.025	.020	.017	.014	.011	.0074	.0038
400	.034	.033	.032	.031	.028	.025	.022	.017	.015	.012	.0094	.0064	.0033
500	.030	.030	.029	.027	.025	.023	.019	.015	.013	.011	.0084	.0057	.0029
600	.028	.027	.026	.025	.023	.021	.018	.014	.012	.0099	.0076	.0052	.0027
750	.025	.024	.024	.022	.021	.018	.016	.013	.011	.0089	.0068	.0047	.0024
1000	.021	.021	.020	.019	.018	.016	.014	.011	.0093	.0077	.0059	.0041	.0021

TABLE VII (*Pages 299-305*)

TABLE OF THE SQUARES AND SQUARE ROOTS OF THE NUMBERS FROM  
1 TO 1000

Taken from Barlow's Tables, in which square roots are given to the seventh decimal place and cubes, cube roots, and reciprocals are also given for numbers up to 10,000.

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
1	1	1.000	51	26 01	7.141	101	1 02 01	10.050
2	4	1.414	52	27 04	7.211	102	1 04 04	10.100
3	9	1.732	53	28 09	7.280	103	1 06 09	10.149
4	16	2.000	54	29 16	7.348	104	1 08 16	10.198
5	25	2.236	55	30 25	7.416	105	1 10 25	10.247
6	36	2.449	56	31 36	7.483	106	1 12 36	10.296
7	49	2.646	57	32 49	7.550	107	1 14 49	10.344
8	64	2.828	58	33 64	7.616	108	1 16 64	10.392
9	81	3.000	59	34 81	7.681	109	1 18 81	10.440
10	1 00	3.162	60	36 00	7.746	110	1 21 00	10.488
11	1 21	3.317	61	37 21	7.810	111	1 23 21	10.536
12	1 44	3.464	62	38 44	7.874	112	1 25 44	10.583
13	1 69	3.606	63	39 69	7.937	113	1 27 69	10.630
14	1 96	3.742	64	40 96	8.000	114	1 29 96	10.677
15	2 25	3.873	65	42 25	8.062	115	1 32 25	10.724
16	2 56	4.000	66	43 56	8.124	116	1 34 56	10.770
17	2 89	4.123	67	44 89	8.185	117	1 36 89	10.817
18	3 24	4.243	68	46 24	8.246	118	1 39 24	10.863
19	3 61	4.359	69	47 61	8.307	119	1 41 61	10.909
20	4 00	4.472	70	49 00	8.367	120	1 44 00	10.954
21	4 41	4.583	71	50 41	8.426	121	1 46 41	11.000
22	4 84	4.690	72	51 84	8.484	122	1 48 84	11.045
23	5 29	4.796	73	53 29	8.544	123	1 51 29	11.091
24	5 76	4.899	74	54 76	8.602	124	1 53 76	11.136
25	6 25	5.000	75	56 25	8.660	125	1 56 25	11.180
26	6 76	5.099	76	57 76	8.718	126	1 58 76	11.225
27	7 29	5.196	77	59 29	8.775	127	1 61 29	11.269
28	7 84	5.292	78	60 84	8.832	128	1 63 84	11.314
29	8 41	5.385	79	62 41	8.888	129	1 66 41	11.358
30	9 00	5.477	80	64 00	8.944	130	1 69 00	11.402
31	9 61	5.568	81	65 61	9.000	131	1 71 61	11.446
32	10 24	5.657	82	67 24	9.055	132	1 74 24	11.489
33	10 89	5.745	83	68 89	9.110	133	1 76 89	11.533
34	11 56	5.831	84	70 56	9.165	134	1 79 56	11.576
35	12 25	5.916	85	72 25	9.220	135	1 82 25	11.619
36	12 96	6.000	86	73 96	9.274	136	1 84 96	11.662
37	13 69	6.083	87	75 69	9.327	137	1 87 69	11.705
38	14 44	6.164	88	77 44	9.381	138	1 90 44	11.747
39	15 21	6.245	89	79 21	9.434	139	1 93 21	11.790
40	16 00	6.325	90	81 00	9.487	140	1 96 00	11.832
41	16 81	6.403	91	82 81	9.539	141	1 98 81	11.874
42	17 64	6.481	92	84 64	9.592	142	2 01 64	11.916
43	18 49	6.557	93	86 49	9.644	143	2 04 49	11.958
44	19 36	6.633	94	88 36	9.695	144	2 07 36	12.000
45	20 25	6.708	95	90 25	9.747	145	2 10 25	12.042
46	21 16	6.782	96	92 16	9.798	146	2 13 16	12.083
47	22 09	6.856	97	94 09	9.849	147	2 16 09	12.124
48	23 04	6.928	98	96 04	9.899	148	2 19 04	12.166
49	24 01	7.000	99	98 01	9.950	149	2 22 01	12.207
50	25 00	7.071	100	1 00 00	10.000	150	2 25 00	12.247

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
151	2 28 01	12.288	201	4 04 01	14.177	251	6 30 01	15.843
152	2 31 04	12.329	202	4 08 04	14.213	252	6 35 04	15.875
153	2 34 09	12.369	203	4 12 09	14.248	253	6 40 09	15.906
154	2 37 16	12.410	204	4 16 16	14.283	254	6 45 16	15.937
155	2 40 25	12.450	205	4 20 25	14.318	255	6 50 25	15.969
156	2 43 36	12.490	206	4 24 36	14.353	256	6 55 36	16.000
157	2 46 49	12.530	207	4 28 49	14.387	257	6 60 49	16.031
158	2 49 64	12.570	208	4 32 64	14.422	258	6 65 64	16.062
159	2 52 81	12.610	209	4 36 81	14.457	259	6 70 81	16.093
160	2 56 00	12.649	210	4 41 00	14.491	260	6 76 00	16.125
161	2 59 21	12.689	211	4 45 21	14.526	261	6 81 21	16.155
162	2 62 44	12.728	212	4 49 44	14.560	262	6 86 44	16.186
163	2 65 69	12.767	213	4 53 69	14.595	263	6 91 69	16.217
164	2 68 96	12.806	214	4 57 96	14.629	264	6 96 96	16.248
165	2 72 25	12.845	215	4 62 25	14.663	265	7 02 25	16.279
166	2 75 56	12.884	216	4 66 56	14.697	266	7 07 56	16.310
167	2 78 89	12.923	217	4 70 89	14.731	267	7 12 89	16.340
168	2 82 24	12.961	218	4 75 24	14.765	268	7 18 24	16.371
169	2 85 61	13.000	219	4 79 61	14.799	269	7 23 61	16.401
170	2 89 00	13.038	220	4 84 00	14.832	270	7 29 00	16.432
171	2 92 41	13.077	221	4 88 41	14.866	271	7 34 41	16.462
172	2 95 84	13.115	222	4 92 84	14.900	272	7 39 84	16.492
173	2 99 29	13.153	223	4 97 29	14.933	273	7 45 29	16.523
174	3 02 76	13.191	224	5 01 76	14.967	274	7 50 76	16.553
175	3 06 25	13.229	225	5 06 25	15.000	275	7 56 25	16.583
176	3 09 76	13.266	226	5 10 76	15.033	276	7 61 76	16.613
177	3 13 29	13.304	227	5 15 29	15.067	277	7 67 29	16.643
178	3 16 84	13.342	228	5 19 84	15.100	278	7 72 84	16.673
179	3 20 41	13.379	229	5 24 41	15.133	279	7 78 41	16.703
180	3 24 00	13.416	230	5 29 00	15.166	280	7 84 00	16.733
181	3 27 61	13.454	231	5 33 61	15.199	281	7 89 61	16.763
182	3 31 24	13.491	232	5 38 24	15.232	282	7 95 24	16.793
183	3 34 89	13.528	233	5 42 89	15.264	283	8 00 89	16.823
184	3 38 56	13.565	234	5 47 56	15.297	284	8 06 56	16.852
185	3 42 25	13.601	235	5 52 25	15.330	285	8 12 25	16.882
186	3 45 96	13.638	236	5 56 96	15.362	286	8 17 96	16.912
187	3 49 69	13.675	237	5 61 69	15.395	287	8 23 69	16.941
188	3 53 44	13.711	238	5 66 44	15.427	288	8 29 44	16.971
189	3 57 21	13.748	239	5 71 21	15.460	289	8 35 21	17.000
190	3 61 00	13.784	240	5 76 00	15.492	290	8 41 00	17.029
191	3 64 81	13.820	241	5 80 81	15.524	291	8 46 81	17.059
192	3 68 64	13.856	242	5 85 64	15.556	292	8 52 64	17.088
193	3 72 49	13.892	243	5 90 49	15.588	293	8 58 49	17.117
194	3 76 36	13.928	244	5 95 36	15.620	294	8 64 36	17.146
195	3 80 25	13.964	245	6 00 25	15.652	295	8 70 25	17.176
196	3 84 16	14.000	246	6 05 16	15.684	296	8 76 16	17.205
197	3 88 09	14.036	247	6 10 09	15.716	297	8 82 09	17.234
198	3 92 04	14.071	248	6 15 04	15.748	298	8 88 04	17.263
199	3 96 01	14.107	249	6 20 01	15.780	299	8 94 01	17.292
200	4 00 00	14.142	250	6 25 00	15.811	300	9 00 00	17.321

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
301	9 06 01	17.349	351	12 32 01	18.735	401	16 08 01	20.025
302	9 12 04	17.378	352	12 39 04	18.762	402	16 16 04	20.050
303	9 18 09	17.407	353	12 46 09	18.788	403	16 24 09	20.075
304	9 24 16	17.436	354	12 53 16	18.815	404	16 32 16	20.100
305	9 30 25	17.464	355	12 60 25	18.841	405	16 40 25	20.125
306	9 36 36	17.493	356	12 67 36	18.868	406	16 48 36	20.149
307	9 42 49	17.521	357	12 74 49	18.894	407	16 56 49	20.174
308	9 48 64	17.550	358	12 81 64	18.921	408	16 64 64	20.199
309	9 54 81	17.578	359	12 88 81	18.947	409	16 72 81	20.224
310	9 61 00	17.607	360	12 96 00	18.974	410	16 81 00	20.248
311	9 67 21	17.635	361	13 03 21	19.000	411	16 89 21	20.273
312	9 73 44	17.664	362	13 10 44	19.026	412	16 97 44	20.298
313	9 79 69	17.692	363	13 17 69	19.053	413	17 05 69	20.322
314	9 85 96	17.720	364	13 24 96	19.079	414	17 13 96	20.347
315	9 92 25	17.748	365	13 32 25	19.105	415	17 22 25	20.372
316	9 98 56	17.776	366	13 39 56	19.131	416	17 30 56	20.396
317	10 04 89	17.804	367	13 46 89	19.157	417	17 38 89	20.421
318	10 11 24	17.833	368	13 54 24	19.183	418	17 47 24	20.445
319	10 17 61	17.861	369	13 61 61	19.209	419	17 55 61	20.469
320	10 24 00	17.889	370	13 69 00	19.235	420	17 64 00	20.494
321	10 30 41	17.916	371	13 76 41	19.261	421	17 72 41	20.518
322	10 36 84	17.944	372	13 83 84	19.287	422	17 80 84	20.543
323	10 43 29	17.972	373	13 91 29	19.313	423	17 89 29	20.567
324	10 49 76	18.000	374	13 98 76	19.339	424	17 97 76	20.591
325	10 56 25	18.028	375	14 06 25	19.365	425	18 06 25	20.616
326	10 62 76	18.055	376	14 13 76	19.391	426	18 14 76	20.640
327	10 69 29	18.083	377	14 21 29	19.416	427	18 23 29	20.664
328	10 75 84	18.111	378	14 28 84	19.442	428	18 31 84	20.688
329	10 82 41	18.138	379	14 36 41	19.468	429	18 40 41	20.712
330	10 89 00	18.166	380	14 44 00	19.494	430	18 49 00	20.736
331	10 95 61	18.193	381	14 51 61	19.519	431	18 57 61	20.761
332	11 02 24	18.221	382	14 59 24	19.545	432	18 66 24	20.785
333	11 08 89	18.248	383	14 66 89	19.570	433	18 74 89	20.809
334	11 15 56	18.276	384	14 74 56	19.596	434	18 83 56	20.833
335	11 22 25	18.303	385	14 82 25	19.621	435	18 92 25	20.857
336	11 28 96	18.330	386	14 89 96	19.647	436	19 00 96	20.881
337	11 35 69	18.358	387	14 97 69	19.672	437	19 09 69	20.905
338	11 42 44	18.385	388	15 05 44	19.698	438	19 18 44	20.928
339	11 49 21	18.412	389	15 13 21	19.723	439	19 27 21	20.952
340	11 56 00	18.439	390	15 21 00	19.748	440	19 36 00	20.976
341	11 62 81	18.466	391	15 28 81	19.774	441	19 44 81	21.000
342	11 69 64	18.493	392	15 36 64	19.799	442	19 53 64	21.024
343	11 76 49	18.520	393	15 44 49	19.824	443	19 62 49	21.048
344	11 83 36	18.547	394	15 52 36	19.849	444	19 71 36	21.071
345	11 90 25	18.574	395	15 60 25	19.875	445	19 80 25	21.095
346	11 97 16	18.601	396	15 68 16	19.900	446	19 89 16	21.119
347	12 04 09	18.628	397	15 76 09	19.925	447	19 98 09	21.142
348	12 11 04	18.655	398	15 84 04	19.950	448	20 07 04	21.166
349	12 18 01	18.682	399	15 92 01	19.975	449	20 16 01	21.190
350	12 25 00	18.708	400	16 00 00	20.000	450	20 25 00	21.213

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
451	20 34 01	21.237	501	25 10 01	22.383	551	30 36 01	23.473
452	20 43 04	21.260	502	25 20 04	22.405	552	30 47 04	23.495
453	20 52 09	21.284	503	25 30 09	22.428	553	30 58 09	23.516
454	20 61 16	21.307	504	25 40 16	22.450	554	30 69 16	23.537
455	20 70 25	21.331	505	25 50 25	22.472	555	30 80 25	23.558
456	20 79 36	21.354	506	25 60 36	22.494	556	30 91 36	23.580
457	20 88 49	21.378	507	25 70 49	22.517	557	31 02 49	23.601
458	20 97 64	21.401	508	25 80 64	22.539	558	31 13 64	23.622
459	21 06 81	21.424	509	25 90 81	22.561	559	31 24 81	23.643
460	21 16 00	21.448	510	26 01 00	22.583	560	31 36 00	23.664
461	21 25 21	21.471	511	26 11 21	22.605	561	31 47 21	23.685
462	21 34 44	21.494	512	26 21 44	22.627	562	31 58 44	23.707
463	21 43 69	21.517	513	26 31 69	22.650	563	31 69 69	23.728
464	21 52 96	21.541	514	26 41 96	22.672	564	31 80 96	23.749
465	21 62 25	21.564	515	26 52 25	22.694	565	31 92 25	23.770
466	21 71 56	21.587	516	26 62 56	22.716	566	32 03 56	23.791
467	21 80 89	21.610	517	26 72 89	22.738	567	32 14 89	23.812
468	21 90 24	21.633	518	26 83 24	22.760	568	32 26 24	23.833
469	21 99 61	21.656	519	26 93 61	22.782	569	32 37 61	23.854
470	22 09 00	21.679	520	27 04 00	22.804	570	32 49 00	23.875
471	22 18 41	21.703	521	27 14 41	22.825	571	32 60 41	23.896
472	22 27 84	21.726	522	27 24 84	22.847	572	32 71 84	23.917
473	22 37 29	21.749	523	27 35 29	22.869	573	32 83 29	23.937
474	22 46 76	21.772	524	27 45 76	22.891	574	32 94 76	23.958
475	22 56 25	21.794	525	27 56 25	22.913	575	33 06 25	23.979
476	22 65 76	21.817	526	27 66 76	22.935	576	33 17 76	24.000
477	22 75 29	21.840	527	27 77 29	22.956	577	33 29 29	24.021
478	22 84 84	21.863	528	27 87 84	22.978	578	33 40 84	24.042
479	22 94 41	21.886	529	27 98 41	23.000	579	33 52 41	24.062
480	23 04 00	21.909	530	28 09 00	23.022	580	33 64 00	24.083
481	23 13 61	21.932	531	28 19 61	23.043	581	33 75 61	24.104
482	23 23 24	21.954	532	28 30 24	23.065	582	33 87 24	24.125
483	23 32 89	21.977	533	28 40 89	23.087	583	33 98 89	24.145
484	23 42 56	22.000	534	28 51 56	23.108	584	34 10 56	24.166
485	23 52 25	22.023	535	28 62 25	23.130	585	34 22 25	24.187
486	23 61 96	22.045	536	28 72 96	23.152	586	34 33 96	24.207
487	23 71 69	22.068	537	28 83 69	23.173	587	34 45 69	24.228
488	23 81 44	22.091	538	28 94 44	23.195	588	34 57 44	24.249
489	23 91 21	22.113	539	29 05 21	23.216	589	34 69 21	24.269
490	24 01 00	22.136	540	29 16 00	23.238	590	34 81 00	24.290
491	24 10 81	22.159	541	29 26 81	23.259	591	34 92 81	24.310
492	24 20 64	22.181	542	29 37 64	23.281	592	35 04 64	24.331
493	24 30 49	22.204	543	29 48 49	23.302	593	35 16 49	24.352
494	24 40 36	22.226	544	29 59 36	23.324	594	35 28 36	24.372
495	24 50 25	22.249	545	29 70 25	23.345	595	35 40 25	24.393
496	24 60 16	22.271	546	29 81 16	23.367	596	35 52 16	24.413
497	24 70 09	22.293	547	29 92 09	23.388	597	35 64 09	24.434
498	24 80 04	22.316	548	30 03 04	23.409	598	35 76 04	24.454
499	24 90 01	22.338	549	30 14 01	23.431	599	35 88 01	24.474
500	25 00 00	22.361	550	30 25 00	23.452	600	36 00 00	24.495

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
601	36 12 01	24.515	651	42 38 01	25.515	701	49 14 01	26.476
602	36 24 04	24.536	652	42 51 04	25.534	702	49 28 04	26.495
603	36 36 09	24.556	653	42 64 09	25.554	703	49 42 09	26.514
604	36 48 16	24.576	654	42 77 16	25.573	704	49 56 16	26.533
605	36 60 25	24.597	655	42 90 25	25.593	705	49 70 25	26.552
606	36 72 36	24.617	656	43 03 36	25.612	706	49 84 36	26.571
607	36 84 49	24.637	657	43 16 49	25.632	707	49 98 49	26.589
608	36 96 64	24.658	658	43 29 64	25.652	708	50 12 64	26.608
609	37 08 81	24.678	659	43 42 81	25.671	709	50 26 81	26.627
610	37 21 00	24.698	660	43 56 00	25.690	710	50 41 00	26.646
611	37 33 21	24.718	661	43 69 21	25.710	711	50 55 21	26.665
612	37 45 44	24.739	662	43 82 44	25.729	712	50 69 44	26.683
613	37 57 69	24.759	663	43 95 69	25.749	713	50 83 69	26.702
614	37 69 96	24.779	664	44 08 96	25.768	714	50 97 96	26.721
615	37 82 25	24.799	665	44 22 25	25.788	715	51 12 25	26.739
616	37 94 56	24.819	666	44 35 56	25.807	716	51 26 56	26.758
617	38 06 89	24.839	667	44 48 89	25.826	717	51 40 89	26.777
618	38 19 24	24.860	668	44 62 24	25.846	718	51 55 24	26.796
619	38 31 61	24.880	669	44 75 61	25.865	719	51 69 61	26.814
620	38 44 00	24.900	670	44 89 00	25.884	720	51 84 00	26.833
621	38 56 41	24.920	671	45 02 41	25.904	721	51 98 41	26.851
622	38 68 84	24.940	672	45 15 84	25.923	722	52 12 84	26.870
623	38 81 29	24.960	673	45 29 29	25.942	723	52 27 29	26.889
624	38 93 76	24.980	674	45 42 76	25.962	724	52 41 76	26.907
625	39 06 25	25.000	675	45 56 25	25.981	725	52 56 25	26.926
626	39 18 76	25.020	676	45 69 76	26.000	726	52 70 76	26.944
627	39 31 29	25.040	677	45 83 29	26.019	727	52 85 29	26.963
628	39 43 84	25.060	678	45 96 84	26.038	728	52 99 84	26.981
629	39 56 41	25.080	679	46 10 41	26.058	729	53 14 41	27.000
630	39 69 00	25.100	680	46 24 00	26.077	730	53 29 00	27.019
631	39 81 61	25.120	681	46 37 61	26.096	731	53 43 61	27.037
632	39 94 24	25.140	682	46 51 24	26.115	732	53 58 24	27.055
633	40 06 89	25.159	683	46 64 89	26.134	733	53 72 89	27.074
634	40 19 56	25.179	684	46 78 56	26.153	734	53 87 56	27.092
635	40 32 25	25.199	685	46 92 25	26.173	735	54 02 25	27.111
636	40 44 96	25.219	686	47 05 96	26.192	736	54 16 96	27.129
637	40 57 69	25.239	687	47 19 69	26.211	737	54 31 69	27.148
638	40 70 44	25.259	688	47 33 44	26.230	738	54 46 44	27.166
639	40 83 21	25.278	689	47 47 21	26.249	739	54 61 21	27.185
640	40 96 00	25.298	690	47 61 00	26.268	740	54 76 00	27.203
641	41 08 81	25.318	691	47 74 81	26.287	741	54 90 81	27.221
642	41 21 64	25.338	692	47 88 64	26.306	742	55 05 64	27.240
643	41 34 49	25.357	693	48 02 49	26.325	743	55 20 49	27.258
644	41 47 36	25.377	694	48 16 36	26.344	744	55 35 36	27.276
645	41 60 25	25.397	695	48 30 25	26.363	745	55 50 25	27.295
646	41 73 16	25.417	696	48 44 16	26.382	746	55 65 16	27.313
647	41 86 09	25.436	697	48 58 09	26.401	747	55 80 09	27.331
648	41 99 04	25.456	698	48 72 04	26.420	748	55 95 04	27.350
649	42 12 01	25.475	699	48 86 01	26.439	749	56 10 01	27.368
650	42 25 00	25.495	700	49 00 00	26.458	750	56 25 00	27.386

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
751	56 40 01	27.404	801	64 16 01	28.302	851	72 42 01	29.172
752	56 55 04	27.423	802	64 32 04	28.320	852	72 59 04	29.189
753	56 70 09	27.441	803	64 48 09	28.337	853	72 76 09	29.206
754	56 85 16	27.459	804	64 64 16	28.355	854	72 93 16	29.223
755	57 00 25	27.477	805	64 80 25	28.373	855	73 10 25	29.240
756	57 15 36	27.495	806	64 96 36	28.390	856	73 27 36	29.257
757	57 30 49	27.514	807	65 12 49	28.408	857	73 44 49	29.275
758	57 45 64	27.532	808	65 28 64	28.425	858	73 61 64	29.292
759	57 60 81	27.550	809	65 44 81	28.443	859	73 78 81	29.309
760	57 76 00	27.568	810	65 61 00	28.460	860	73 96 00	29.326
761	57 91 21	27.586	811	65 77 21	28.478	861	74 13 21	29.343
762	58 06 44	27.604	812	65 93 44	28.496	862	74 30 44	29.360
763	58 21 69	27.622	813	66 09 69	28.513	863	74 47 69	29.377
764	58 36 96	27.641	814	66 25 96	28.531	864	74 64 96	29.394
765	58 52 25	27.659	815	66 42 25	28.548	865	74 82 25	29.411
766	58 67 56	27.677	816	66 58 56	28.566	866	74 99 56	29.428
767	58 82 89	27.695	817	66 74 89	28.583	867	75 16 89	29.445
768	58 98 24	27.713	818	66 91 24	28.601	868	75 34 24	29.462
769	59 13 61	27.731	819	67 07 61	28.618	869	75 51 61	29.479
770	59 29 00	27.749	820	67 24 00	28.636	870	75 69 00	29.496
771	59 44 41	27.767	821	67 40 41	28.653	871	75 86 41	29.513
772	59 59 84	27.785	822	67 56 84	28.671	872	76 03 84	29.530
773	59 75 29	27.803	823	67 73 29	28.688	873	76 21 29	29.547
774	59 90 76	27.821	824	67 89 76	28.705	874	76 38 76	29.563
775	60 06 25	27.839	825	68 06 25	28.723	875	76 56 25	29.580
776	60 21 76	27.857	826	68 22 76	28.740	876	76 73 76	29.597
777	60 37 29	27.875	827	68 39 29	28.758	877	76 91 29	29.614
778	60 52 84	27.893	828	68 55 84	28.775	878	77 08 84	29.631
779	60 68 41	27.911	829	68 72 41	28.792	879	77 26 41	29.648
780	60 84 00	27.928	830	68 89 00	28.810	880	77 44 00	29.665
781	60 99 61	27.946	831	69 05 61	28.827	881	77 61 61	29.682
782	61 15 24	27.964	832	69 22 24	28.844	882	77 79 24	29.698
783	61 30 89	27.982	833	69 38 89	28.862	883	77 96 89	29.715
784	61 46 56	28.000	834	69 55 56	28.879	884	78 14 56	29.732
785	61 62 25	28.018	835	69 72 25	28.896	885	78 32 25	29.749
786	61 77 96	28.036	836	69 88 96	28.914	886	78 49 96	29.766
787	61 93 69	28.054	837	70 05 69	28.931	887	78 67 69	29.783
788	62 09 44	28.071	838	70 22 44	28.948	888	78 85 44	29.799
789	62 25 21	28.089	839	70 39 21	28.965	889	79 03 21	29.816
790	62 41 00	28.107	840	70 56 00	28.983	890	79 21 00	29.833
791	62 56 81	28.125	841	70 72 81	29.000	891	79 38 81	29.850
792	62 72 64	28.142	842	70 89 64	29.017	892	79 56 64	29.866
793	62 88 49	28.160	843	71 06 49	29.034	893	79 74 49	29.883
794	63 04 36	28.178	844	71 23 36	29.052	894	79 92 36	29.900
795	63 20 25	28.196	845	71 40 25	29.069	895	80 10 25	29.917
796	63 36 16	28.213	846	71 57 16	29.086	896	80 28 16	29.933
797	63 52 09	28.231	847	71 74 09	29.103	897	80 46 09	29.950
798	63 68 04	28.249	848	71 91 04	29.120	898	80 64 04	29.967
799	63 84 01	28.267	849	72 08 01	29.138	899	80 82 01	29.983
800	64 00 00	28.284	850	72 25 00	29.155	900	81 00 00	30.000

NUM.	SQUARE	Sq. ROOT	NUM.	SQUARE	Sq. ROOT
901	81 18 01	30.017	951	90 44 01	30.838
902	81 36 04	30.033	952	90 63 04	30.854
903	81 54 09	30.050	953	90 82 09	30.871
904	81 72 16	30.067	954	91 01 16	30.887
905	81 90 25	30.083	955	91 20 25	30.903
906	82 08 36	30.100	956	91 39 36	30.919
907	82 26 49	30.116	957	91 58 49	30.935
908	82 44 64	30.133	958	91 77 64	30.952
909	82 62 81	30.150	959	91 96 81	30.968
910	82 81 00	30.166	960	92 16 00	30.984
911	82 99 21	30.183	961	92 35 21	31.000
912	83 17 44	30.199	962	92 54 44	31.016
913	83 35 69	30.216	963	92 73 69	31.032
914	83 53 96	30.232	964	92 92 96	31.048
915	83 72 25	30.249	965	93 12 25	31.064
916	83 90 56	30.265	966	93 31 56	31.081
917	84 08 89	30.282	967	93 50 89	31.097
918	84 27 24	30.299	968	93 70 24	31.113
919	84 45 61	30.315	969	93 89 61	31.129
920	84 64 00	30.332	970	94 09 00	31.145
921	84 82 41	30.348	971	94 28 41	31.161
922	85 00 84	30.364	972	94 47 84	31.177
923	85 19 29	30.381	973	94 67 29	31.193
924	85 37 76	30.397	974	94 86 76	31.209
925	85 56 25	30.414	975	95 06 25	31.225
926	85 74 76	30.430	976	95 25 76	31.241
927	85 93 29	30.447	977	95 45 29	31.257
928	86 11 84	30.463	978	95 64 84	31.273
929	86 30 41	30.480	979	95 84 41	31.289
930	86 49 00	30.496	980	96 04 00	31.305
931	86 67 61	30.512	981	96 23 61	31.321
932	86 86 24	30.529	982	96 43 24	31.337
933	87 04 89	30.545	983	96 62 89	31.353
934	87 23 56	30.561	984	96 82 56	31.369
935	87 42 25	30.578	985	97 02 25	31.385
936	87 60 96	30.594	986	97 21 96	31.401
937	87 79 69	30.610	987	97 41 69	31.417
938	87 98 44	30.627	988	97 61 44	31.432
939	88 17 21	30.643	989	97 81 21	31.448
940	88 36 00	30.659	990	98 01 00	31.464
941	88 54 81	30.676	991	98 20 81	31.480
942	88 73 64	30.692	992	98 40 64	31.496
943	88 92 49	30.708	993	98 60 49	31.512
944	89 11 36	30.725	994	98 80 36	31.528
945	89 30 25	30.741	995	99 00 25	31.544
946	89 49 16	30.757	996	99 20 16	31.559
947	89 68 09	30.773	997	99 40 09	31.575
948	89 87 04	30.790	998	99 60 04	31.591
949	90 06 01	30.806	999	99 80 01	31.607
950	90 25 00	30.822	1000	100 0000	31.623

## APPENDIX THREE

### SUPPLEMENTARY PRACTICE MATERIAL

For additional practice in statistical method Tables VIII to XVI (pages 307-316) are provided. These may be used for practice in —

- (1) finding a mean of distribution ;
- (2) finding a median ;
- (3) finding an average deviation ;
- (4) finding a median deviation ;
- (5) finding a standard deviation ;
- (6) drawing a percentile curve ;
- (7) drawing a histogram ;
- (8) finding the median, upper- and lower-quartile values, and various percentile values such as the 90-percentile and 10-percentile values, in a distribution, the overlapping of two distributions, etc., by means of percentile curves ;
- (9) finding the correspondence between two tests by means of percentile curves and a line of relation ;
- (10) finding the average of a pupil's scores in two or more tests when the scores of all the tests have been transmuted into terms of a single test ;
- (11) finding the average between scores and school marks ;
- (12) finding the coefficient of correlation between two tests ;
- (13) finding the probable error of a coefficient of correlation ;
- (14) finding the reliability of a test ;
- (15) finding the practice effect in a test ;
- (16) finding the (partial) correlation between two variables with the effect of a third eliminated ;
- (17) finding the (multiple) correlation between the combination of two tests and a third, etc.

The medians, coefficients, etc., are quoted as printed with the tables. Any slight differences between your results and these may be due to the use of different methods of calculation.

TABLE VIII<sup>1</sup>

DISTRIBUTION OF SCORES FOR GRADES 2 AND 3 ON EACH OF SIX TESTS

SCORE	DEARBORN		DETROIT		HAGGERTY		MYERS		OTIS		PRESSEY	
	2	3	2	3	2	3	2	3	2	3	2	3
150		1										
145		3										
140		2										
135		4										
130		2										
125	2	1										
120	1	1										
115	2	6										
110	2	2										
105	3	3										
100	1	1										
95	4	6										
90	3	2										
85	2	2				1						3
80	2	1									5	2
75	6	3				1			2		5	7
70	1	1				3			4		7	6
65		2				3					6	6
60	6				2	6	1		2		6	5
55	2				3	6	2	2	4		3	3
50			1			2			5	5		1
45		3		4	1	6	1		8	7	5	6
40		11	13	6	6	5	2	6	8	6	1	2
35		17	22	9	4	4	3	1	5	6		1
30		6	5	6	3	6	7	9	3	5		
25					5	2	6	10		2		1
20					3		6	7	1		1	
15			1		3	1	8	4				
10					1	1	4	2				
5-9						1	1					
Total . .	37	43	39	44	39	44	37	43	39	43	39	43
Median . .	88.7	111.3	38.7	38.9	35.8	49.6	24.6	29.3	42.2	46.8	67.9	67.1
Avg. Dev.	17.6	20.2	4.1	3.3	9.3	13.8	7.6	8.6	7.7	10.7	10.0	11.7
St. Dev. .	20.6	23.8	5.74	4.0	12.13	17.02	8.9	11.38	9.4	13.5	13.2	14.4

<sup>1</sup> From an article entitled "A Study of Intelligence Scales for Grades Two and Three," by J. Cayce Morrison, W. B. Cornell, and Ethel Cornell, in *Journal of Educational Research* for January, 1924, page 47.

TABLE IX.<sup>1</sup> BINET MENTAL AGE, GROUP INTELLIGENCE TEST SCORES, AND HIGH SCHOOL RECORDS OF THE PUPILS IN THE CLASS OF JUNE, 1920

PUPIL	CHRONOLOGICAL AGE	MENTAL AGE	INTELLIGENCE QUOTIENT	OTIS SCORE	AVERAGE RATINGS		CREDITS EARNED IN H. S.
					VIII-A	High School	
1	11-4	17-10	157	159	90	92	4
2	14-1	18-9	133	156	89	85	4
3	12-6	15-8	125	153	88	90	4
4	13-2	16	122	169	82	88	4
5	13-2	15-9	120	145	91	95	4
6	14-6	16-9	116	155	86	88	4
7	14-3	16-5	115	123	89	80	4½
8	13	14-8	113	158	90	91	4
9	13-7	15-3	112	147	81	82	4
10	13-3	14-10	112	159	79	89	4
11	13-9	15-1	110	166	89	88	4
12	14	15-4	110	150	92	84	4
13	13-6	14-8	109	127	85	87	4
14	14-2	15-4	108	127	82	74	2
15	14-9	15-2	103	154	88	79	3
16	15-5	15-7	101	152	85	87	4
17	14-9	14-10	101	103	84	85	4
18	13-9	13-6	98	147	92	89	4
19	15-3	14-8	96	146	93	88	4
20	13-10	13-4	96	136	92	82	4
21	13-7	13-1	96	118	85	78	3
22	15-6	14-10	95	135	87	79	4
23	14-10	14-1	95	133	81	82	4½
24	14-6	13-9	93	124	86	87	4
25	16	14-7	91	120	87	69	1
26	14-6	13-1	89	96	83	82	4
27	14-3	12-5	87	99	87	88	4
28	15-10	13-6	86	127	86	78	3
29	15-2	11-5	75	84	84	80	3½
30	16-9	11-11	74	84	64	87	3½

<sup>1</sup> From an article entitled "A Study of the Use of the Stanford Revision of the Binet-Simon Test as a Guide to Selection of High School Courses," by Sara E. Weisman in *Journal of Educational Research*, Feb., 1923, p. 138.

TABLE X<sup>1</sup>

TOTAL SCORES BY EACH STUDENT AT SUCCESSIVE TRIALS OF THE STONE  
REASONING TEST

STUDENT'S NUMBER	FIRST TRIAL	SECOND TRIAL	THIRD TRIAL	STUDENT'S NUMBER	FIRST TRIAL	SECOND TRIAL	THIRD TRIAL
1	80	60	53	26	54	76	70
2	73	57	55	27	54	54	54
3	70	80	60	28	53	55	63
4	70	69	65	29	52	45	50
5	68	65	73	30	52	40	40
6	67	65	58	31	51	58	60
7	65	67	65	32	51	60	50
8	64	73	66	33	50	60	62
9	63	52	40	34	50	58	74
10	63	40	38	35	48	64	62
11	62	50	65	36	48	44	57
12	61	74	60	37	48	44	35
13	61	70	70	38	47	58	63
14	61	55	70	39	47	45	50
15	61	53	67	40	46	60	56
16	61	49	35	41	46	45	55
17	60	70	60	42	45	57	55
18	60	56	44	43	43	40	60
19	60	55	55	44	42	47	54
20	60	53	48	45	40	48	58
21	58	50	60	46	40	40	50
22	58	50	58	47	40	40	49
23	55	55	53	48	36	40	40
24	55	44	48	49	29	30	42
25	55	30	40				

<sup>1</sup> From an article entitled "Reducing the Variability of Teachers' Marks," by E. J. Ashbaugh in *Journal of Educational Research* for March, 1924, page 188.

TABLE XI<sup>1</sup>

## DISTRIBUTION OF HENMON VOCABULARY SCORES

SCORE	NUMBER OF SEMESTERS OF LATIN STUDIED							
	1	2	3	4	5	6	7	8
48-50 . . .		35	13	204	24	165	9	81
44-47 . . .		254	82	656	94	352	27	121
40-43 . . .	2	478	145	669	37	184	11	40
36-39 . . .	4	624	127	400	17	48	1	18
32-35 . . .	3	616	69	229	11	10	1	6
28-31 . . .	5	511	31	117	4	9		
24-27 . . .	5	296	11	65		1		
20-23 . . .	3	185	8	30	2			
16-19 . . .	5	84	1	14				
12-15 . . .	10	50	1	3				
8-11 . . .	2	15		1				
4-7 . . .	2	4						
0-3 . . .		5						
Total Number of Pupils .	41	3,157	488	2,388	189	769	49	266
Medians . .	22.7	34.8	39.9	42.0	45.0	45.5	45.8	46.3
Lower Quartile	14.4	29.2	36.0	37.4	41.4	42.7	43.6	44.1
Upper Quartile	31.2	39.9	43.3	45.6	47.0	47.7	47.6	48.7
Q . . . . .	8.4	5.4	3.6	4.2	2.8	2.5	2.0	2.3

<sup>1</sup> From an article entitled "The Status of Certain Basic Latin Skills," by Leo J. Brueckner in *Journal of Educational Research* for May, 1924, page 395.

TABLE XII<sup>1</sup>

## DISTRIBUTION OF HENMON SENTENCE TEST SCORES

(Scored on Unit-Credit Basis)

SCORES (Sentences Correct)	NUMBER OF SEMESTERS OF LATIN STUDIED							
	1	2	3	4	5	6	7	8
10 . . . . .				3	1	1		1
9 . . . . .				10	2	8	2	5
8 . . . . .		4	5	33	8	22	3	16
7 . . . . .		14	8	112	14	45	1	23
6 . . . . .		61	19	245	25	103	9	37
5 . . . . .	1	169	58	419	33	169	13	64
4 . . . . .	4	369	106	490	47	180	8	63
3 . . . . .	3	743	135	485	38	148	9	31
2 . . . . .	9	1,197	114	380	16	77	4	20
1 . . . . .	6	474	32	154	5	13		4
0 . . . . .	18	126	11	57		3		2
Total Number of Pupils .	41	3,157	488	2,388	189	769	49	266
Medians . . .	1.5	2.8	3.6	4.2	4.8	4.8	5.3	5.2
Lower Quartile	0.6	2.2	2.7	3.0	3.7	3.6	2.9	4.1
Upper Quartile	2.8	3.8	4.7	5.5	6.1	5.9	6.3	6.4
Q . . . . .	1.1	.8	1.0	1.3	1.2	1.1	1.2	1.2

<sup>1</sup>From an article entitled "The Status of Certain Basic Latin Skills," by Leo J. Brueckner in *Journal of Educational Research* for May, 1924, page 395.

TABLE XIII.<sup>1</sup> SCORES AND ESTIMATED INTELLIGENCE OF SIXTY-FOUR HIGH SCHOOL PUPILS IN SIX TESTS<sup>2</sup>

SUBJECT	OTIS	B-S. M.A.	ALPHA	MILLER	TERMAN	LEARNING TEST	AGE	EST. INTEL.
1	203	216	172	114	184	8.3	176	8.5
2	144	205	126	76	149	7	189	6.3
3	127	195	108	63	129	15	209	5
4	140	195	101	62	131	23.3	204	5
5	177	210	147	90	162	-3.7	191	6.7
6	92	184	89	42	96	1.7	244	1.7
7	151	189	104	81	148	10.6	175	4.7
8	135	173	82	57	115	14.6	230	4
9	102	193	97	61	102	7.7	216	3.5
10	106	180	79	29	77	10	209	2.7
11	154	195	106	67	123	14.7	215	7
12	113	179	67	41	89	7	202	3.7
13	162	211	143	93	167	12	187	5.7
14	144	188	123	70	140	9	212	5.2
15	144	189	114	45	122	15.7	192	5
16	144	202	105	74	130	8	210	6.2
17	171	210	136	84	169	7.3	233	6.5
18	181	219	177	95	175	40.6	187	9
19	119	202	83	60	118	7.7	224	5.2
20	194	220	153	96	173	26	181	7
21	131	184	81	69	114	19.3	200	5.6
22	194	201	139	80	151	11	166	7.8
23	134	180	105	58	89	6.4	228	4.7
24	145	190	104	70	119	7.6	162	7.6
25	163	214	123	64	122	15.4	178	5.2
26	155	205	132	68	124	12.4	201	6.3
27	129	193	103	39	147	10	200	5.5
28	147	211	114	66	123	23	222	5.2
29	124	190	92	56	107	5.67	257	5.5
30	183	192	125	84	174	16.3	199	6

<sup>1</sup> These data were kindly furnished by Dr. A. M. Jordan, University of North Carolina, Chapel Hill, North Carolina.

<sup>2</sup> Otis = Otis Group Intelligence Scale, Advanced Examination. B-S. M.A. = Mental age of Stanford Revision of Binet-Simon Tests. Alpha = Army Alpha. Miller = Miller Mental Ability Test. Terman = Terman Group Test of Mental Ability. Learning Test (See *Journal of Educational Psychology*, September, 1923). Age = Chronological age. Estimated Intelligence = Average estimate of four mature critic teachers.

TABLE XIII (Continued)

SUB- JECT	OTIS	B-S. M.A.	ALPHA	MILLER	TERMAN	LEARNING TEST	AGE	EST. INTEL.
31	181	196	120	85	153	5.7	217	4.7
32	145	188	119	64	106	15.6	232	6
33	206	222	169	90	195	16.3	217	7.5
34	200	205	157	102	178	19	161	9.5
35	152	210	130	87	156	5.7	199	6
36	160	228	122	79	136	13	187	8
37	171	210	147	87	162	10.67	234	4.2
38	176	206	139	73	143	20.3	214	4.5
39	128	187	122	54	123	15	235	5.3
40	146	201	111	65	105	12.3	210	5
41	189	211	146	90	158	19	228	8.5
42	110	156	72	58	109	10.6	216	3
43	100	179	72	63	101	12.7	223	4.2
44	116	180	106	50	116	12.7	199	4.2
45	160	219	112	78	157	9.6	228	5
46	194	193	128	93	155	10.4	171	7
47	108	202	98	54	110	29.4	224	3
48	164	184	138	83	102	8	197	5.6
49	149	173	121	75	90	20.6	207	3.5
50	171	211	151	61	124	11	225	6.7
51	111	160	80	59	87	7	192	2.7
52	156	190	123	54	102	12	198	4
53	158	199	133	70	109	17.4	197	3.2
54	148	217	151	81	145	11	206	4.7
55	106	149	107	57	80	4.3	221	4.5
56	168	187	125	70	119	11.7	166	4.6
57	136	189	109	74	128	15.7	227	4.5
58	189	228	156	94	168	15.6	206	8.2
59	177	210	130	88	164	16.4	189	6.7
60	159	183	128	86	153	13.7	218	6.5
61	133	201	115	60	118	9.4	252	5.6
62	145	209	111	44	88	15.6	185	5.3
63	108	164	97	43	61	3.3	204	3
64	151	206	124	64	94	10	195	3.3

(For Intercorrelations see next page.)

TABLE XIII (Continued). INTERCORRELATIONS

		II	III	IV	V	VI	VII	VIII	S.D.
Otis . . . . .	I	.65	.84	.82	.79	.27	-.43	.72	28.4
B-S. M. A. . . .	II		.68	.55	.67	.27	-.05	.59	17.2
Alpha . . . . .	III			.77	.73	.21	-.28	.65	24.7
Miller . . . . .	IV				.81	.17	-.32	.68	17.6
Terman . . . . .	V					.21	-.24	.66	29.9
Learning Test . .	VI						.12	.26	6.6
Age . . . . .	VII							-.31	21.6
Est. Intel. . . .	VIII								1.7

TABLE XIV<sup>1</sup>

CORRELATION BETWEEN THE I.Q.'S OF ARMY GROUP EXAMINATION *a*  
AND THE QUALITY OF SCHOOL WORK OF 494 HIGH SCHOOL PUPILS

SCHOOL MARKS	ARMY GROUP EXAMINATION <i>a</i> I.Q.'s										TOTALS
	84 or Lower	85-89	90-94	95-99	100-104	105-109 ( <i>Median</i> )	110-114	115-119	120-124	125 or Above	
90 or above . .	—	—	3	2	2	3	6	6	6	6	34
85-89 . . . .	—	—	5	9	18	24	22	24	8	4	114
80-84 ( <i>Median</i> )	1	2	9	28	41	46	30	13	9	3	182
75-79 . . . .	1	3	9	19	19	19	19	7	2	—	98
70-74 . . . .	2	3	4	14	11	11	7	1	1	—	54
65-69 . . . .	—	1	1	4	1	1	2	1	—	—	11
60-64 . . . .	—	—	—	—	—	1	—	—	—	—	1
Totals . . . .	4	9	31	76	92	105	86	52	26	13	494

Medians: I.Q., 106; school marks, 82

Semi-interquartile ranges: I.Q.'s, 6 points; school marks, 4½ per cent

Correlation, Pearson's formula, 0.343; P.E., 0.027

<sup>1</sup> Tables XIV and XV are from an article entitled "Psychological Tests as a Means of Measuring the Probable School Success of High School Pupils," by W. M. Proctor in *Journal of Educational Research* for April, 1920, page 259.

TABLE XV

CORRELATION BETWEEN THE I.Q.'S OF THE ARMY GROUP EXAMINATION  
ALPHA AND THE QUALITY OF SCHOOL WORK OF 480 HIGH SCHOOL  
PUPILS

SCHOOL MARKS	ARMY GROUP EXAMINATION ALPHA I.Q.'s										TOTALS
	84 or Lower	85-89	90-94	95-99	100-104	105-109 (Median)	110-114	115-119	120-124	125 or Above	
90 or over . . .	—	—	—	3	3	15	12	9	9	5	56
85-89 . . . .	—	—	—	8	17	15	24	13	6	6	89
80-84 . . . .	—	—	4	6	22	21	20	10	5	1	89
75-79 (Median)	—	—	7	25	33	23	10	7	4	—	109
70-74 . . . .	—	4	10	18	14	22	12	1	1	—	82
65-69 . . . .	1	3	3	12	7	8	8	1	—	—	43
60-64 . . . .	—	—	2	5	3	1	1	—	—	—	12
Totals . . . .	1	7	26	77	99	105	87	41	25	12	480

Medians: I.Q., 106; school marks, 79

Semi-interquartile ranges: I.Q.'s, 6 points; school marks,  $6\frac{1}{2}$  per cent

Correlation, Pearson's formula, 0.413; P.E., 0.026

TABLE XVI.<sup>1</sup> INTELLIGENCE QUOTIENTS OBTAINED FROM TERMAN AND STANFORD-BINET EXAMINATIONS OF 155 CHILDREN

No.	INTELLIGENCE QUOTIENTS		No.	INTELLIGENCE QUOTIENTS		No.	INTELLIGENCE QUOTIENTS		No.	INTELLIGENCE QUOTIENTS	
	Terman	Stanford Binet		Terman	Stanford Binet		Terman	Stanford Binet		Terman	Stanford Binet
1	152	139	40	114	109	79	104	87	118	95	100
2	148	120	41	114	104	80	104	102	119	95	95
3	143	132	42	113	114	81	104	100	120	95	90
4	140	132	43	113	86	82	103	98	121	94	95
5	136	116	44	113	110	83	103	100	122	94	79
6	135	112	45	112	98	84	103	90	123	94	77
7	132	121	46	112	100	85	103	94	124	94	98
8	132	104	47	111	108	86	102	105	125	93	67
9	132	119	48	110	103	87	102	89	126	93	85
10	131	118	49	110	83	88	102	93	127	93	88
11	130	114	50	110	103	89	102	97	128	92	87
12	130	124	51	109	100	90	102	80	129	92	96
13	127	108	52	109	112	91	102	97	130	92	81
14	127	114	53	109	118	92	101	105	131	92	90
15	126	111	54	109	87	93	101	100	132	91	78
16	126	125	55	108	87	94	100	96	133	91	79
17	126	115	56	108	104	95	100	94	134	91	84
18	126	120	57	107	97	96	100	84	135	91	88
19	125	117	58	107	92	97	100	90	136	90	88
20	124	110	59	107	105	98	100	85	137	90	81
21	123	115	60	107	103	99	100	100	138	90	72
22	123	112	61	106	87	100	100	70	139	90	83
23	121	124	62	106	98	101	99	99	140	90	101
24	121	111	63	106	94	102	99	94	141	90	85
25	121	124	64	106	88	103	99	86	142	89	78
26	120	128	65	105	99	104	99	92	143	89	87
27	119	119	66	105	88	105	98	104	144	89	78
28	118	113	67	105	87	106	98	96	145	89	92
29	118	109	68	105	102	107	98	111	146	89	88
30	118	100	69	105	96	108	98	89	147	89	78
31	117	104	70	105	109	109	98	91	148	88	81
32	116	104	71	105	94	110	98	89	149	87	97
33	115	95	72	104	98	111	97	77	150	87	76
34	115	107	73	104	96	112	97	102	151	85	78
35	115	109	74	104	86	113	97	91	152	85	71
36	115	107	75	104	102	114	97	90	153	84	81
37	115	93	76	104	119	115	96	84	154	82	81
38	115	98	77	104	102	116	95	92	155	81	58
39	114	96	78	104	101	117	95	82			

$r = .80$ , P.E. = .019.

<sup>1</sup> A slight modification of a table from an article entitled "A Study of the Binet and Terman Intelligence Tests with Eleven-Year-Old Children," by George T. Avery in *Journal of Educational Research* for May, 1923, page 432.

## APPENDIX FOUR

### ANSWERS TO EXERCISES

Ex. 1. 89.

Ex. 2. You should get the same result.

Ex. 3. You should get the same result.

Ex. 4. (a) 12.9, (b) 25.6, (c) 57.8.

Ex. 5.

GRADE	SCORE																		NUM.	MED.
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18				
6A	1	2	1	2	3	3	3	5	2	8	4	7	2	0	0	1	44	11		
6B	2	1	1	1	0	6	2	7	8	6	4	2	2	1	0	1	44	11		
7A			3	2	1	4	3	4	8	3	6	3	4				41	11		
7B	1			1	1	2	3	8	2	3	6	6	2	1			36	11		

GRADE	SCORE																	
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
6A	1	2½	4	5½	8	11	14	18	21½	26½	32½	38	42½				44	
6B	1½	3	4	5		8½	12½	17	24½	31½	36½	39½	41½	43			44	

Ex. 8. A score of 10 has a higher rank in Grade 6A.

Ex. 9.

SCORE	5	6	7	8	9	10	11	12	13	14	15	16
P.R.	0	7	14	25	38	50	66	80	90	97		100

Ex. 14. Frequencies beginning in the interval 20 to 24: 3, 2, 5, 4, 3, 10, 7, 5, 2, 4, ending in the interval 65 to 69. Number of cases, 45.

Ex. 15.	$1 \times 10 = 10$	
	$2 \times 9 = 18$	$173 \div 38 = 4.55.$
	$2 \times 8 = 16$	Auxiliary value, 4 = Score of 22.
	$3 \times 7 = 21$	.55 of 5 points = about 3 points.
	$4 \times 6 = 24$	(The exact mean of the actual score
	$5 \times 5 = 25$	cannot be found, of course, since we are
	$6 \times 4 = 24$	using mid-points only.)
	$8 \times 3 = 24$	Mean = $22 + 3 = 25.$
	$4 \times 2 = 8$	
	$3 \times 1 = 3$	
	<hr/>	
	38	173

Ex. 18.	GRADE	10%ILE	25%ILE	50%ILE	75%ILE	90%ILE
	5A	19	27	37	47	57
	6A	28	36	47	54	63

- Ex. 20. 5A, I.Q.R. = about 20 points (27 to 47).  
 5A, 10-90-percentile range = about 37 points (19 to 57).  
 6A, I.Q.R. = about 17 points (37, by smooth curve, to 54).  
 6A, 10-90-percentile range = about 34 points (28 to 62 by curve).

- Ex. 21. About 25 per cent of 6A fall below median of 5A.  
 About 25 per cent of 5A exceed median of 6A.

- Ex. 22. Med. Dev. of 5A = 2 points; of 5B, 2 points.

- Ex. 23. Avg. Dev. of 5A = 2.1 points; of 5B, 2.0 points.

- Ex. 25.  $.80 \times 43 = 34.4$  = most probable value of A.D.  
 $.67 \times 43 = 29$  = most probable value of M.D.  
 $2 \times 29 = 58$  = most probable value of I.Q.R.

- Ex. 26.  $24 \div 2 = 12$  = most probable value of M.D.  
 $1.18 \times 12 = 14.2$  = most probable value of A.D.  
 $1.48 \times 12 = 17.8$  = most probable value of  $\sigma$ .

- Ex. 27. Pupils 2, 3, 4, 5, and 11 did better in the N.I.T.  
 (Those pupils did better in the N.I.T. whose points lie above — *i.e.*, to the left of — the line)

Ex. 30.	I.E. SCORES . . . . .	8	16	24	32	40
	CORRESPONDING N.I.T. SCORES	56	70	85	98	111

Ex. 31.

N.I.T.	I.E.
42	0
43	1
44	2
45	2
46	3
47	3
48	4
49	4
50	5

Ex. 32. 2, 1, 7, 8, 5, 6, 3, 9, 10, 4.

Ex. 33. 2, 1, 7, 5, 8 and 9, 6, 3, 10, 4.

Ex. 34. (14 yrs.) 49; (53) 15 yrs.; ( $9\frac{1}{2}$  yrs.) 19; (10 yrs. 3 mos.) 25; (11 yrs. 5 mos.) 34; (16 yrs.) 56; (17 yrs.) 58; (18 yrs.) 59; (19 yrs.) 59; (20 yrs.) 59; (21) 9 yrs. 9 mos.; (37) 11 yrs. 10 mos.; (46) 13 yrs. 4 mos.; (MA) 13 yrs. 4 mos.; (47) 13 yrs. 7 mos.; (51) 14 yrs. 6 mos.; (57) 16 yrs. 5 mos.; (58) 17 yrs.; (59) 18 yrs.

Ex. 35. See paragraph following the exercise.

Ex. 36. The means are, from left to right, for the second diagram, 7.5, 7.75, 8, 8.25 and 8.5, and for the fourth diagram, 6.5, 7.25, 8, 8.75, and 9.5.

Ex. 37. .50.

Ex. 38. See Figure 54.

Ex. 39. See Figure 52.

Ex. 40. Coef. of cor. =  $60 \pm .04$ .Ex. 41. Coef. of cor. =  $.56 \pm .04$ .

Ex. 42. .905.

Ex. 43. See Col. 4 of the following table.

Ex. 44. See Cols. 7 and 10 of the following table.

TABLE OF ANSWERS TO EXERCISES 43 AND 44<sup>1</sup>

VARIABLES	<i>M</i>	<i>S</i>	<i>r</i>	P.E. <sub><i>r</i></sub>	$\Sigma G$	<i>R</i>	<i>r</i>	$\Sigma D^2$	$\rho$	<i>r</i>
O.A.T. — N.I.T.	309	434	.712	.065	53.5	.524	.757	769	.737	.753
O.A.T. — T.G.T.	312	398	.782	.051	46	.591	.819	553	.811	.824
O.A.T. — O.I.E.	392	474	.827	.042	50.5	.551	.783	630	.785	.799
S.A.T. — N.I.T.	252	345	.730	.061	54	.520	.753	692	.764	.779
S.A.T. — T.G.T.	256	317	.808	.046	39	.653	.871	437	.851	.862
S.A.T. — O.I.E.	306	377	.812	.045	51.5	.542	.774	644.5	.780	.794
N.I.T. — T.G.T.	222	283	.784	.051	44.5	.605	.832	478.5	.836	.848
N.I.T. — O.I.E.	250	337	.742	.059	47.5	.578	.808	561.5	.808	.821
T.G.T. — O.I.E.	272	309	.880	.030	41	.636	.855	434.5	.851	.862

Ex. 46. .98, .92, .87, .60, .44; .98.

Ex. 47. .81, .95.

Ex. 49. .97.

Ex. 49.  $r_{12.3} = .20, .33, .47, .67, .08, .22, .41, .88, .69, .46$ .

Ex. 50. Ratio of  $\sigma$ 's = 12.91 : 10.

$$\text{New coef.} = 1 - \left[ \left( \frac{10}{12.91} \right)^2 \times (1 - .50) \right] = .70.$$

Ex. 51.  $R_{C.12} = .58, .69, .81, .78, .76, .50, .36, .74, .78, .72$ .

Ex. 52. See fifth column of the table of answers to Exercise 43.

Ex. 53. .75, .82, .86, .82, .88, .90; .89, .92, .94.

<sup>1</sup> In obtaining these coefficients, scores were grouped as suggested in the exercise. If other groupings are used, slightly varying coefficients may result. Calculations were made with an 18-inch slide rule. Variations of a point or two in the thousandths place are to be expected.

## APPENDIX FIVE

### BIBLIOGRAPHY

#### Textbooks dealing primarily with statistical methods

- Barlow's *Tables of Squares, Cubes, Square Roots, Cube Roots, and Reciprocals*. E. and F. N. Spon, Ltd., London. 200 pages. 1921. (An excellent reference book for finding squares, square roots, etc., on numbers from 1 to 10,000.)
- BOWLEY, A. L. *Elements of Statistics*. P. S. King & Son, London; Charles Scribner's Sons, New York. 1907.
- BROWN, W. M., and THOMSON, GODFREY H. *The Essentials of Mental Measurement*. Cambridge University Press, Cambridge, England. 216 pages. 1920. (A revision and extension of Brown's *Essentials of Mental Measurement*.)
- DAVENPORT, C. B. *Statistical Methods*. John Wiley & Sons, New York. 1904. (Especially adapted to the study of biological variation. Summarizes all mathematical formulæ, with brief descriptions of methods. Complete and useful tables, containing probability integrals, Beta and Gamma functions, logarithms, etc.)
- ELDERTON, W. P. and E. M. *Primer of Statistics*. A. & C. Black, London. 1910.
- JONES, D. CARADOG. *A First Course in Statistics*. G. Bell & Sons, Ltd., London. 286 pages. 1921. (Deals somewhat at length with curve fitting.)
- KELLEY, TRUMAN L. *Statistical Method*. American Book Company, New York. 390 pages. 1923. (A very technical treatise on statistical method involving calculus for use by advanced students. Contains an extensive bibliography of technical articles and a 6-place table of values of the variables of a normal surface of frequency.)
- "Tables: To Facilitate the Calculation of Partial Coefficients of Correlation and Regression Coefficients." University of Texas Bulletin No. 27. Austin, Texas, 1916. (Out of print.)
- KING, W. I. *The Elements of Statistical Method*. The Macmillan Company, New York. 1915.
- MERRIMAN, M. *Method of Least Squares*. John Wiley & Sons, New York. 1894.
- RIETZ, H. L., et al. *Handbook of Mathematical Statistics*. Houghton Mifflin Company, Boston. 221 pages. 1924. (A highly technical symposium on statistical methods.)
- RUGG, HAROLD O. *Statistical Methods Applied to Education*. Houghton Mifflin Company, Boston. 410 pages. 1917.
- SECRIST, HORACE. *An Introduction to Statistical Methods*. The Macmillan Company, New York. 1917.

- THORNDIKE, EDWARD L. *Mental and Social Measurements*. Teachers College, Columbia University, New York. 277 pages. 1904. (A fairly elementary treatise.)
- THURSTONE, L. L. *The Fundamentals of Statistics*. The Macmillan Company, New York. 1925.
- YULE, G. UDNY. *An Introduction to the Theory of Statistics*. J. B. Lippincott Company, Philadelphia. 382 pages. Third edition, 1916. (One of the early textbooks on statistical method, giving the first description of the method of partial correlation.)
- ZIZEK, F. *Statistical Averages*. (Translated by W. M. Persons.) Henry Holt & Co., New York. 1913.

#### Textbooks dealing incidentally with statistical methods

- GILLILAND, A. R., and JORDAN, R. H. *Educational Measurements and the Classroom Teacher*. The Century Company, New York. 269 pages. 1924. Chapter XVIII, "Statistical and Graphic Methods," 243-265.
- HINES, HARLAN C. *A Guide to Educational Measurements*. Houghton Mifflin Company, Boston. 270 pages. 1923. Part I, "Statistical Methods," 3-48.
- MCCALL, WILLIAM A. *How to Measure in Education*. The Macmillan Company, New York. 416 pages. 1922. Part III, "Tabular, Graphic, and Statistical Methods," 321-408.
- *How to Experiment in Education*. The Macmillan Company, New York. 281 pages. 1923.
- MONROE, WALTER S. *The Theory of Educational Measurement*. Houghton Mifflin Company, Boston. 364 pages. 1923. Chapter XII, "Statistical Methods: The Frequency of Distribution and Its Description," 296-335.
- MONROE, WALTER S., DE VOSS, JAMES C., and KELLEY, F. J. *Educational Tests and Measurements*. Houghton Mifflin Company, Boston. 309 pages. 1917. Chapter VIII, "Statistical Methods."
- PRESSEY, SIDNEY L. and LUELLA COLE. *Introduction to the Use of Standard Tests*. World Book Company, Yonkers-on-Hudson, New York. 263 pages. 1923. Chapter III, "The Common Sense of Statistics," 36-59.
- TRABUE, MARION R. *Measuring Results in Education*. American Book Company, New York. 492 pages. 1924. Chapters IX, X, XI, XVII.
- WHIPPLE, G. M. *Manual of Mental and Physical Tests*. Warwick & York, Inc., Baltimore. Second edition, 1914. Vol. I, "Simpler Processes," Chapter III. (A brief summary of the statistical treatment of measures as applied to problems of educational psychology.)
- WILSON, G. M., and HOKE, KREMER J. *How to Measure*. The Macmillan Company, New York. 285 pages. 1920. Chapter XI, "Statistical Terms and Methods."

## Articles dealing primarily or incidentally with statistical method

To assist the reader in finding the articles bearing on certain topics, the following list is given, each topic being followed by the numbers of the chief articles bearing on that subject.

Bibliography: 39.

Central tendency (median, mean, mode): 10, 47.

Correlation: 2, 3, 5, 29, 30, 37, 49, 53, 56, 57, 58, 63, 65, 66, 70, 73, 76, 80, 81, 82, 84, 90.

Correspondence between and combining of measures, and transmutation: 24, 25, 27, 36, 38, 53, 55, 89.

Distributions: 68, 69.

Intelligence *vs.* achievement: 15, 42.

Partial and multiple correlations, regression equations and prediction: 4, 8, 9, 30, 33, 34, 35, 36, 54, 77.

Percentile curves and percentile ranks: 59, 75.

Rating: 12, 14, 51, 64, 67, 72.

Reliability: 1, 7, 15, 17, 18, 23, 26, 31, 32, 41, 42, 43, 44, 50, 52, 58, 61, 65, 71, 72, 84, 88.

Standardization of terminology: 13, 45.

Suggested studies: 60.

T-scale: 48, 87.

Validity: 39.

Variability (Median deviation, average deviation, standard deviation, etc.): 11, 19, 28, 30, 46, 79.

Weighting of items in tests: 21, 86.

1. ASKER, WILLIAM. "The Reliability of Tests Requiring Alternative Responses." *Journal of Educational Research*, Vol. 9 (1924), pages 234-240. (Contains interesting data on probability.)
2. AYRES, LEONARD P. "The Application of Tables of Distribution of a Shorter Method for Computing Coefficients of Correlation." *Journal of Educational Research*, Vol. 1 (1920), pages 295-298.
3. ——— "A Shorter Method for Computing the Coefficient of Correlation." *Journal of Educational Research*, Vol. 1 (1920), pages 216-221.
4. ——— "Coefficients of Regression." *Journal of Educational Research*, Vol. 1 (1920), pages 398-402.
5. ——— "Substituting Small Numbers for Large Ones in the Computation of Coefficients of Correlation." *Journal of Educational Research*, Vol. 2 (1920), pages 502-504.
6. BAKER, H. J. "A Short Method of Determining Probable Error." *Journal of Educational Research*, Vol. 2 (1920), pages 592-593. (Gives a chart for finding the probable error of a coefficient of correlation in terms of the coefficient itself. Short.)
7. BATSON, WILLIAM H. "Reliability of the True-False Form of Ex-

- amination." *Educational Administration and Supervision*, February, 1924, pages 95-102.
8. BUCKINGHAM, B. R. "Multiple Correlation." *Journal of Educational Research*, Vol. 9 (1924), pages 149-154. (Editorial.)
  9. ——— "Partial Correlation." *Journal of Educational Research*, Vol. 7 (1923), pages 344-349. (Editorial.)
  10. ——— "The Use of the Median in Supervision." *Journal of Educational Research*, Vol. 5 (1922), pages 154-159. (Editorial.)
  11. ——— "Variability." *Journal of Educational Research*, Vol. 4 (1921), pages 221-226. (Editorial.)
  12. ——— "A Human Ladder." *Journal of Educational Research*, Vol. 4 (1921), pages 143-146. (Editorial. Discusses rating scales.)
  13. BUCKINGHAM, MCCALL, OTIS, RUGG, TRABUE, COURTIS. "Report of the Standardization Committee." *Journal of Educational Research*, Vol. 4 (1921), pages 78-80. (Discusses standardization of terminology.)
  14. BYRNE, LEE. "A Method of Equalizing the Rating of Teachers." *Journal of Educational Research*, Vol. 4 (1921), pages 102-108.
  15. CHAPMAN, J. CROSBY. "The Unreliability of the Difference between Intelligence and Educational Ratings." *Journal of Educational Psychology*, Vol. 14 (1923), pages 103-108.
  16. ——— "Some Elementary Statistical Considerations in Educational Measurements." *Journal of Educational Research*, Vol. 4 (1921), pages 212-220.
  17. COBB, MARGARET V. "One Element in the Probable Error of a Mental Age Measurement." *Journal of Educational Psychology*, Vol. 13 (1922), pages 236-240.
  18. CRUM, W. L. "Note on the Reliability of a Test, with Special Reference to the Examinations Set by the College Entrance Board." *The American Mathematical Monthly*, Vol. 30 (1923), September-October.
  19. DAVIS, STURGISS B. "A Modification of the Short Method for Finding the Standard Deviation of a Distribution of Measures." (Short.) *Journal of Educational Research*, Vol. 8 (1923), pages 183-185.
  20. DOLL, EDGAR A. "Improper Use of the I.Q." *Journal of Delinquency*, Vol. 5 (1920), pages 67-70.
  21. DOUGLAS, HARL R., and SPENCER, PETER L. "Is It Necessary to Weight Exercises in Standard Tests?" *Journal of Educational Psychology*, Vol. 14 (1923), pages 109-112.
  22. FRANZEN, RAYMOND. "Two Notes on Statistical Method." *Journal of Educational Psychology*, Vol. 15 (1924), pages 250-251.
  23. ——— "Statistical Issues: I. Commensuration and Age; II. The Reliability of the Difference between Intelligence and Educational Ratings; III. The Determination of Reliability." *Journal of Educational Psychology*, Vol. 15 (1924), pages 367-382.
  24. GARRETT, HENRY E. "An Empirical Study of the Various Methods of Combining Incomplete Order of Merit Ratings." *Journal of Educational Psychology*, Vol. 15 (1924), pages 157-171.

25. GATES, ARTHUR I. "The Relation of Quality and Speed of Performance: A Formula for Combining the Two in the Case of Handwriting." *Journal of Educational Psychology*, Vol. 15 (1924), pages 129-144.
26. GEYER, DENTON L. "The Reliability of Rankings by Group Intelligence Tests." *Journal of Educational Psychology*, Vol. 13 (1922), pages 43-49.
27. GOODRICH, T. V. "Putting Test Scores into Form to Facilitate Comparison." *Journal of Educational Research*, Vol. 8 (1923), pages 377-388.
28. HENMON, V. A. C. "Comparative Variability at Different Ages." *Journal of Educational Psychology*, Vol. 13 (1922), pages 17-29.
29. HOLZINGER, KARL J. "Formulas for the Correlation between Ratios." *Journal of Educational Psychology*, Vol. 14 (1923), pages 344-347.
30. — "Charlier's Control Checks on the Computation of the Standard Deviation and Coefficients of Correlation and Regression." *Journal of Educational Research*, Vol. 8 (1923), pages 275-277. (Short.)
31. — "Inferences from Small Samples and Differences." *Journal of Educational Research*, Vol. 6 (1922), pages 175-176. (Discusses reliability of a difference. Short.)
32. — "On the Assumption that Errors of Estimate Are Equal in Narrow and Wide Ranges." *Journal of Educational Research*, Vol. 4 (1921), pages 237-239. (Short.)
33. HUFFAKER, O. L. "A Contribution to the Technique of Partial Correlations." *Journal of Applied Psychology*, Vol. 6 (1923), pages 135-142.
34. HULL, CLARK L. "A Device for Determining Coefficients of Partial Correlation." *Psychological Review*, Vol. 28 (1921), pages 377-383.
35. — "Prediction Formulæ for Teams of Aptitude Tests." *Journal of Applied Psychology*, Vol. 7 (1923), pages 277-284.
36. — "The Joint Yield from Teams of Tests." *Journal of Educational Psychology*, Vol. 14 (1923), pages 396-406.
37. — "The Computation of Pearson's  $r$  from Ranked Data." *Journal of Applied Psychology*, Vol. 6 (1922), pages 385-390.
38. — "The Conversion of Test Scores into Series Which Shall Have Any Assigned Mean and Degree of Dispersion." *Journal of Applied Psychology*, Vol. 6 (1922), pages 298-300.
39. JORDAN, A. M. "The Validation of Intelligence Tests." *Journal of Educational Psychology*, Vol. 14 (1923), pages 348-365, 414-428. (Contains a long bibliography on correlations.)
40. KELLEY, TRUMAN L. "Chart to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations." *Stanford University Publications, School of Education, Special Monograph No. 1* (1921).
41. — "Further Note upon the Probable Error of the Mean." *Journal of Educational Psychology*, Vol. 14 (1923), pages 497-498.
42. — "A New Method for Determining the Significance of Differ-

- ences in Intelligence and Achievement Scores." *Journal of Educational Psychology*, Vol. 14 (1923), pages 321-333.
43. KELLEY, TRUMAN L. "Note on the Reliability of a Test: A Reply to Dr. Crum's Criticism." *Journal of Educational Psychology*, Vol. 15 (1924), pages 193-204.
  44. — "Reliability of Test Scores." *Journal of Educational Research*, Vol. 3 (1921), pages 370-379.
  45. — "Report of the Subcommittee on Statistical Methods of the Standardization Committee (Tentative)." *Journal of Educational Research*, Vol. 4 (1921), pages 77-78.
  46. — "The Measurement of Overlapping." *Journal of Educational Psychology*, Vol. 10 (1919), pages 458-461. (Shows how apparent overlapping of distributions of scores of different groups is greater than the actual overlapping because of the unreliability of the tests.)
  47. MCCALL, WILLIAM A. "How to Compute the Median." *Teachers College Record*, Vol. 21 (1920), pages 124-138.
  48. — "Proposed Uniform Method of Scale Construction: With Application to a New Reading Scale." *Teachers College Record*, Vol. 22 (1921), pages 31-51. (Explains the construction of the T-scale.)
  49. MONROE, WALTER S. "Effect of Grade Distribution upon Coefficients of Correlation." *Journal of Educational Research*, Vol. 2 (1920), page 776.
  50. MORTON, R. L. "The Reliability of Measurements in Spelling." *Journal of Educational Method*, Vol. 3 (1924), pages 321-328.
  51. — "The Normal Distribution of School Marks." *Journal of Educational Research*, Vol. 3 (1921), pages 73-74. (Shows how to give marks of A, B, C, D, and E on the 7%-24%-38%-24%-7% basis.)
  52. NYGAARD, P. H. "The Advantages of the Probable Error of Measurement as a Criterion of the Reliability of a Test or Scale." *Journal of Educational Psychology*, Vol. 14 (1923), pages 407-413.
  53. OTIS, ARTHUR S. "An Absolute Point Scale for the Group Measurement of Intelligence." *Journal of Educational Psychology*, Vol. 9 (1918), pages 239-261, 333-348; Appendix III, "Some Mathematical Reasoning with Regard to Criteria of Tests of Intelligence." (Shows how the coefficient of correlation between any one weighted test and the weighted composite of a number of tests may be found from the weights and the coefficients of intercorrelation between the tests.)
  54. — "The Derivation of Simpler Forms of Regression Equations." *Journal of Educational Psychology*, Vol. 8 (1917), pages 619-622. (Contains a review of Truman L. Kelley's "Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations," which was Bulletin of the University of Texas, May 10, 1916, No. 27 — out of print.)
  55. — "The Method of Finding the Correspondence between Scores in Two Tests." *Journal of Educational Psychology*, Vol. 13 (1922), pages 529-545.

56. OTIS, ARTHUR S. "A Method of Inferring a Change in a Coefficient of Correlation Resulting from a Change in the Heterogeneity of the Group." *Journal of Educational Psychology*, Vol. 13 (1922), pages 293-5.
57. — "The Otis Correlation Chart." *Journal of Educational Research*, Vol. 8 (1923), pages 440-448.
58. — "The Reliability of Spelling Scales, Involving a Deviation Formula for Correlation." *School and Society*, Vol. 4 (1916), Nos. 96-99.
59. — "Some Logical Aspects of the Binet Scale." *Psychological Review*, Vol. 23 (1916), pages 129-152, 165-179. (Shows some early uses of percentile curves and discusses the growth of mental ability.)
60. — "Suggested Studies in the Field of Mental Testing." *Journal of Educational Method*, Vol. 1 (1922), pages 220-232. (A list of eight studies suitable for subjects of masters' theses.)
61. OTIS, ARTHUR S., and KNOLLIN, HERBERT E. "The Reliability of the Binet Scale and Pedagogical Scales." *Journal of Educational Research*, Vol. 4 (1921), pages 121-142. (Discusses a new formula for correlation — the difference formula.)
62. PRESSEY, SIDNEY L. "Suggestions Looking toward a Fundamental Revision of Current Statistical Procedure, as Applied to Tests." *Psychological Review*, Vol. 27 (1920), pages 466-472.
63. RAND, GERTRUDE. "The Use of the Correlation Graph with Half-Sigma Class Intervals." *Journal of Educational Research*, Vol. 9 (1924), pages 213-222.
64. REAM, M. J. "A Statistical Method for Incomplete Order of Merit Ratings." *Journal of Applied Psychology*, Vol. 5 (1921), pages 261-266.
65. SHEN, EUGENE. "The Standard Error of Certain Estimated Coefficients of Correlation." *Journal of Educational Psychology*, Vol. 15 (1924), pages 462-465.
66. SPEARMAN, C. "Correlations of Sums and Differences." *British Journal of Psychology*, Vol. 5, pages 419-426.
67. STENQUIST, JOHN L. "An Improved Form for Rating by the Order of Merit Method." *Journal of Educational Psychology*, Vol. 11 (1920), pages 526-528.
68. — "Plotting Graphic Distribution Charts with a Typewriter." *Journal of Educational Research*, Vol. 8 (1923), pages 468-472. (Short.)
69. SYMONDS, PERCIVAL M. "A Second Approximation to the Curve of Distribution of Intelligence of the Population of the United States, with a Note on the Standardization of the Stanford Revision of the Binet-Simon Scale." *Journal of Educational Psychology*, Vol. 14 (1923), pages 65-82.
70. THOMSON, GODFREY H., and PINTNER, RUDOLF. "Spurious Correlation and Relationship between Tests." *Journal of Educational Psychology*, Vol. 15 (1924), pages 433-444.
71. THORNDIKE, EDWARD L. "The Reliability and Significance of Tests of Intelligence." *Journal of Educational Psychology*, Vol. 11 (1920), pages 284-287.

72. THORNDIKE, EDWARD L. "A Constant Error in Psychological Ratings." *Journal of Applied Psychology*, Vol. 4 (1920), pages 25-29.
73. ——— "The Construction and Interpretation of Correlation Tables." *Journal of Educational Research*, Vol. 7 (1923), pages 199-212.
74. THORNDIKE, EDWARD L., and COURTIS, S. A. "Correction Formula for Addition Tests." *Teachers College Record*, Vol. 21 (1920), pages 1-24.
75. THURSTONE, L. L. "The Calculation and Interpretation of Percentile Ranks." *Journal of Educational Research*, Vol. 6 (1922), pages 225-235.
76. ——— "A Data Sheet for the Pearson Correlation Coefficient." *Journal of Educational Research*, Vol. 6 (1922), pages 49-56.
77. ——— "A Scoring Method for Mental Tests." *Psychological Bulletin*, Vol. 16 (1919).
78. ——— "A Statistical Method for the Treatment of School-Survey Data." *The School Review*, Vol. 25 (1917), pages 322-330.
79. ——— "A Method of Calculating the Pearson Correlation Coefficient without the Use of Deviations." *Psychological Bulletin*, Vol. 14 (1917), pages 28-32.
80. TOOPS, HERBERT A. "Computing Intercorrelations of Tests on the Adding Machine." *Journal of Applied Psychology*, Vol. 6 (1922), pages 172-184.
81. ——— "Eliminating the Pitfalls in Solving Correlation: A Printed Correlation Form." *Journal of Experimental Psychology*, Vol. 4 (1921), pages 434-446.
82. ——— "Solving Intercorrelations by Polar Coördinates." *Journal of Experimental Psychology*, Vol. 5 (1922), pages 68-75.
83. ——— "Calculating the Standard Deviation by Plotting and without Grouping the Raw Measures." *Journal of Experimental Psychology*, Vol. 7 (1924), pages 75-80.
84. TOOPS, HERBERT A., and MINER, ZAIDA F. "A Serviceable P.E.<sub>r</sub> Table." *Journal of Educational Research*, Vol. 9 (1924), pages 63-68. (For finding the number of cases necessary to yield a correlation coefficient of any given value with a P.E.<sub>r</sub> of any given value.)
85. TOOPS, HERBERT A., and PINTNER, RUDOLF. "Curves of Growth of Intelligence." *Journal of Experimental Psychology*, Vol. 3 (1920), pages 231-242.
86. WEST, PAUL V. "The Significance of Weighted Scores." *Journal of Educational Psychology*, Vol. 15 (1924), pages 302-308.
87. WILSON, G. M. "Note Concerning a T-Scale for Wilson Language Error Test." *Journal of Educational Psychology*, Vol. 15 (1924), 118-119.
88. WOOD, BEN D. "The Reliability of Prediction of Proportions on the Basis of Random Sampling." *Journal of Educational Research*, Vol. 4 (1921), pages 390-395.
89. WOODWORTH, R. S. "Combining the Results of Several Tests: A Study in Statistical Method." *Psychological Review*, Vol. 19 (1912).
90. WOOLBERT, CHARLES H. "Tables to Facilitate the Computation of Coefficients of Correlation by the Rank Difference Method." *Journal of Applied Psychology*, Vol. 4 (1920), pages 115-125.

# INDEX

References set in boldface type indicate pages on which definitions or explanatory statements are given of the entries to which they belong.

- Abbreviations, 289
- Accomplishment ratio, 4, **172**, 337
- Age calculator, 333
- Alternative form, 4
- Array, 177
- Ashbaugh, E. J., iv, 309
- Assumed mean, use of, 191
- Attenuation, **227**; correction for, 228
- Average, 6, 7
- Average deviation, 88, **90**, 93
- Avery, George T., iv, 316
  
- Bar graph, 81
- Barlow's Tables, 298
- Binet mental age, 144
- Binet-Simon Tests, 144, 150, 156
- Binominal expansion, 71
- Bonner, H. R., iv, 163
- Bright, normal, and dull sections, 278
- Brightness, 148; coefficient of, 153; Index of, 155
- Brown, William, 246
- Brown's Formula, 263
- Brueckner, Leo J., iv, 310, 311
- "B score," 166
- Buckingham, B. R., iv, 246
  
- Central tendency, 6, 9, 18, **19**
- Clark, John R., iv
- Classification Index, **278**
- Classification within grades, 277
- Class intervals, **37**, 194
- Class ranges, 194
- Coefficient of alienation, 220, 222
- Coefficient of Brightness, **153**
- Coefficient of contingency, 215
- Coefficient of correlation, 181; significance of, 182; in relation to causation, 183; calculation of, 186; probable error of, 256
- Coins, flipping of, 70, 183
- Comparison, of individuals within a group, 20; of groups, 23
- Cornell, Ethel, 307
- Cornell, W. B., 307
- Correction for attenuation, **228**
- Correlation, 4; meaning of, **175**; negative, 200, 201; rank methods of computing, 206; by unlike signs, 212; and prognosis, 217; effect of errors of measurement upon, 227; effect of heterogeneity on, 235. *See also* Coefficient of correlation, Multiple correlation, and Partial correlation
- Correlation chart, 4
- Correlation ratio, 214
- Correspondence, between tests, 101; use of percentile graph in finding, 108
- Criterion, 217, 238
- Cumulative errors, **58**
  
- Deviation, average, 88, **90**, 93; standard, 88, **91**, 93; median, 88, 93, 94, 99, 224, 294
- Difference, reliability of, 260
- Difference method, 189
- Displacement, interpretation of a coefficient of correlation in terms of, 223
- Distribution, **14**; bimodal, **19**; grouped, 36; normal curve of, 74; skewed, **75**
  
- Educational quotient, 171, 337
- Errors, of prediction, 218; of measurement, 226, 255
  
- Fallibility of test scores, 247

- Footrule, 210  
 Form, alternative, 4  
 Franzen, Raymond, 172, 174  
 Frequency, 15; normal surface of, 73, 224, 287
- Galton, Francis, 186  
 Geometric mean, 7  
 Grade norms, 3; how to obtain from age norms, 163  
 Grade status, 165  
 Greek alphabet, 290  
 Grouping, 194; effect of, 64  
 "G score," 166
- Heterogeneity, effect of, on correlation, 235; and reliability, 254  
 Heubner, B. P., iv, 205  
 Higher Examination. *See* Otis Self-Administering Tests of Mental Ability  
 Histogram, 31, 42, 73, 75, 80, 81  
 Homogeneity, 3  
 Hooker, Grover, 66  
 Hudelson English Composition Scale, 168
- Index of Brightness, 155  
 Individual differences, 140  
 Intelligence quotient, 99, 148, 156, 230; invalidity of, 150; from group tests, 154; slide rule, 337  
 Intermediate Examination. *See* Otis Self-Administering Tests of Mental Ability  
 Interpolation, 129, 167, 169, 211, 212  
 Interpretation Chart, 157, 281  
 Interquartile range, 86, 88, 93
- Jordan, A. M., iv, 312
- Kelley, Truman L., iv, 75, 87, 214, 216, 220, 246, 263  
 Knollin, H. E., 216
- Law of probability, 70  
 Line graph, 34, 43  
 Line of diagonals, 34, 60, 81; how to draw on a percentile graph, 55; effect of smoothing, 78  
 Line of relation, 105, 112, 117, 128, 169, 180
- McCall, William A., iv, 122, 166  
 McGaughy, J. Ralph, iv  
 Mean, the, 6, 19; formula for, 7; geometric, 7; short method for finding, 10, 18, 178; finding from a distribution, 17; finding by method of substitution, 39; assumed, 191; probable error of, 261  
 Median, the, 11; formula for finding, 12; by sorting, 13; finding from a distribution, 15; finding from a step graph, 33; of a grouped distribution, 43; finding by estimate, 44; finding by consulting original papers, 44; finding graphically, 45, 47; of a large group, 48; precise statements of, 48; by smoothing, 84  
 Median deviation, 88, 93, 94, 99, 224, 294  
 Median score, 19  
 Mental ability, growth of, 132, 135, 138; definition of, 136  
 Mental age, 143, 337; Binet, 144; true, 144, 146; fictitious, 146  
 Mental maturity, 140  
 Method of excesses, 9  
 Mid-score, 11  
 Mode, 18, 19  
 Morrison, J. Cayce, iv, 307  
 Multiple correlation, 238; formula for, 239  
 Murphy, H. H., iv
- National Intelligence Test, 101, 118, 155

- Negative correlation, 200, **201**  
 Non-linear relationship, **211**, 214, 216  
 Normal distribution, 294; law of, 68, **72**  
 Normal surface of distribution, 73  
 Normal surface of frequency, **73**, 224, 287  
 Norms, 37, **143**; grade, 3, 161; mental age, 160; age, 160  
 Otis Achievement Test. *See* Otis Classification Test  
 Otis Classification Test, 125, 173, 203, 208, 269, 271, 278  
 Otis Correlation Chart, 192  
 Otis Group Intelligence Scale, 153, 155, 160  
 Otis Self-Administering Tests of Mental Ability, 55, 66, 101, 118, 136, 142, 144, 156, 173  
 Overlapping, 3, 15, **87**  
 Partial correlation, 230; nature of, 233  
 Pearson, Karl, 76, 186  
 Percentage table, 59  
 Percentile curve, **80**; advantages of, 81; how to draw, 81  
 Percentile graph, 4, 53; advantages of, 53; general utility of, 54; use of, in averaging scores and teacher's marks, 126  
 Percentile rank, 95, 156, 294; precise rule for finding, 26; approximate method of finding, 27; in terms of variability, 98  
 Percentile ranks, disadvantages of averaging, 118  
 Pintner, Rudolf, iv  
 Practice effect, **264**  
 Probability surface, **75**  
 Probable error, **89**, 248, **249**, 252; of a coefficient of correlation, 256, **258**, 298; of a difference, 260; of a mean, 261; of a standard deviation, 261; of coefficients of partial and multiple correlation, 262  
 Proctor, W. M., iv, 314  
 Product-moment method, 186, 189  
 Prognostic test, 217  
 Quartile scores, 33; finding graphically, 47  
 Range, **85**; of score, 15  
 Rank (rank order), 20, 81; relative, 23; percentile, **24**  
 Rank methods of computing correlation, 206, 295, 296  
 Rating scale, 286  
 Regrading, 270  
 Regression equation, **243**  
 Regression line, 181, 186  
 Reliability, **227**, 247; measures of, 248; and heterogeneity, 254  
 Reliability coefficient, **228**, **248**, 251, 252  
 Rietz, H. L., 246  
 Sampling, 160, 257, 298  
 Scale chart, 60  
 Scatter diagram, **176**, 186  
 Semi-interquartile range, 88  
 Skewness, measurement of, **76**  
 Spearman, 263  
 Spearman Footrule, **210**, 295  
 Square roots, 299  
 Squares, 299  
 Standard deviation, 88, **91**, 93; of a difference, 260; probable error of, 261  
 Stanford Achievement Test, 203, 208  
 Stanford Revision of the Binet-Simon Test, 145, 149  
 Statistical tables, 291  
 Step graph, 43, 81  
 Subject ages, **171**; fictitious, 171  
 Subject ratios, 172

- Symbols, 289  
 Symonds, Percival M., iv, 68, 244  
 Table of correspondence, 106, 115;  
     between teacher's marks and  
     scores, 129  
 Table of Products, 199  
 Teacher's marks, 186; averaging  
     with scores, 124  
 Teaching load, 279  
 10-90-percentile ranges, 87  
 Terman, Lewis M., iv, 98, 149  
 Test scores, validity of, 136  
 Theobald, Jacob, iv  
 Thomson, Godfrey H., iv, 246  
 Three-track plan, 280  
 Thurstone, L. L., 191  
 Transmuting scores, 119  
 True mental age, 144, 146  
 T-score method, disadvantages of,  
     124  
 T scores, 122  
 Universal Percentile Graph, 62  
 Unlike signs, 297; method of, 213,  
     223  
 Validity, 256  
 Variability, 3, 85, 121  
 Webber, Elmer H., iv, 203  
 Weighting of tests, 120, 131, 240  
 Weisman, Sara E., iv, 308  
 Whitney, Frederick L., 246  
 Wood, Ben D., iv  
 Yule, G. Udny, 246

## AGE CALCULATOR

Page 335, when detached, together with the table on page 334, constitute an age calculator by which it is possible to find the age of a pupil in months when the date of birth is known. Ages may be found to the nearest month. Only one setting is needed to find any number of ages on a given date. The calculator is good for 10 years. To prepare for use, detach Sheet 2 (page 335) and cut out the rectangle as directed. The rectangle is thrown away.

To insure against loss of Sheet 2 it might be well to fold over the edge of page 334 and slip Sheet 2 under the fold when not in use.

**Setting.** Let us suppose that we wish to find the ages of a group of pupils on September 15, 1927. The calculator is set as follows: First, look at the right-hand side of the table and find September. Set the loose sheet so that September just shows. (Cover August.) Next move the loose sheet to the right or left so that 1927 just shows on the lower line of the table. (1928 is covered.) If properly set, the top row of numbers visible in the table will be 213, 201, 189, 177, etc., with the 213 directly below 1909.

**Finding ages.** We are now ready to find any pupil's age. Let us suppose that some pupil was born September 15, 1917. To find his age, first look for 1917 in the row of dates on the loose sheet, then run down the column under 1917 to the line opposite September on the loose sheet. You will see that the number in that cell is 120. The pupil, therefore, is 120 months old (just 10 years). The same procedure is used for any other dates.

**Correction.** The age so found is correct within one month. To find the age to the nearest month (i.e., within one-half month), a correction may need to be applied. If the day of the month on which the pupil was born is more than 15 days *earlier* than the present day of the month, *add* 1 month to the age found in the table. Thus, if this is the 20th of the month and the pupil was born before the 5th of the month, add 1 month to the age as found in the table. If the day of the month on which the pupil was born is more than 15 days *later* than the present day of the month, subtract 1 month from the age as found in the table. These corrections will render the age correct within one-half month.

**Converting to years and months.** Note that ages expressed in months can be converted into years and months and vice versa very easily by means of the chronological-age or mental-age scale of the IQ Slide Rule. For example, 170 months = 14 years, 2 months.

# AGE CALCULATOR

SHEET 1

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	↑ Set so present month just shows
205	193	181	169	157	145	133	121	109	97	85	73	61	49
206	194	182	170	158	146	134	122	110	98	86	74	62	50
207	195	183	171	159	147	135	123	111	99	87	75	63	51
208	196	184	172	160	148	136	124	112	100	88	76	64	52
209	197	185	173	161	149	137	125	113	101	89	77	65	53
210	198	186	174	162	150	138	126	114	102	90	78	66	54
211	199	187	175	163	151	139	127	115	103	91	79	67	55
212	200	188	176	164	152	140	128	116	104	92	80	68	56
213	201	189	177	165	153	141	129	117	105	93	81	69	57
214	202	190	178	166	154	142	130	118	106	94	82	70	58
215	203	191	179	167	155	143	131	119	107	95	83	71	59
216	204	192	180	168	156	144	132	120	108	96	84	72	60
217	205	193	181	169	157	145	133	121	109	97	85	73	61
218	206	194	182	170	158	146	134	122	110	98	86	74	62
219	207	195	183	171	159	147	135	123	111	99	87	75	63
220	208	196	184	172	160	148	136	124	112	100	88	76	64
221	209	197	185	173	161	149	137	125	113	101	89	77	65
222	210	198	186	174	162	150	138	126	114	102	90	78	66
223	211	199	187	175	163	151	139	127	115	103	91	79	67
224	212	200	188	176	164	152	140	128	116	104	92	80	68
225	213	201	189	177	165	153	141	129	117	105	93	81	69
226	214	202	190	178	166	154	142	130	118	106	94	82	70
227	215	203	191	179	167	155	143	131	119	107	95	83	71

Set so present year just shows here

1935 1934 1933 1932 1931 1930 1929 1928 1927 1926 1925

Leave this page in the book. Detach the next page (Sheet 2) and cut out the rectangle as directed. See directions for use on page 333.

Cut on this line to detach page

## AGE CALCULATOR

### SHEET 2

1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931

Cut on this line

Dec.  
Nov.  
Oct.  
Sept.  
Aug.  
July  
June  
May  
Apr.  
Mar.  
Feb.  
Jan.

Cut on this line

To prepare for use detach this page and cut out this rectangle



## IQ SLIDE RULE

**To prepare for use.** First cut very accurately on the lines at the top and bottom of the page and throw away the strips of paper. This leaves the scales the proper length. Next cut the full length of the page along the two vertical lines. This will detach a wide strip of paper containing the mental age and the IQ scales and a narrow strip containing nothing. In each case cut on the side of the line in such a way that both lines go with the narrow strip that is thrown away, leaving both the IQ scale and the chronological-age scale with graduations only.

Next fold the attached portion of the page on the line as directed, so that the line is on the outside of the fold.

Then slip the IQ scale under the chronological-age scale and up against the fold so that the chronological-age scale laps slightly over the mental-age scale. The slide rule is then ready for use.

**To find an IQ.** To divide a pupil's mental age (MA) by his chronological age (CA), slide the detached slip to such a position that the point on the mental-age scale representing the MA of the pupil is directly opposite the point on the chronological-age scale representing the CA of the pupil. The point on the IQ scale at the edge of the page represents the IQ of the pupil.

For example, let us suppose that the pupil's chronological age is 10 years and his mental age is 12 years. Find the point on the chronological-age scale representing 10 years and move the slide until the point representing 12 on the mental-age scale is directly opposite it. Notice that the slide is projecting at the top (left) and the point on the IQ scale (on the slide) at the exact edge of the folded page is 120. This is the pupil's IQ. IQ's below 100 are read at the bottom edge of the page. It is customary in finding an IQ to use 16 years as the maximum chronological age; i.e., when a pupil's CA is more than 16, call it 16.

Any other quotients, of course, may be found by this slide rule. Thus, to find an educational quotient by dividing EA by CA, use exactly the same procedure, reading the EA on the mental-age scale.

To find an accomplishment ratio by dividing EA by MA, read the EA on the mental-age scale and the MA on the chronological-age scale. (The dividend is always read on the moving scale and the divisor on the stationary scale.)

**For general dividing purposes.** Note that any number between 36 and 216 may be divided by any other number between 36 and 216 by this slide rule, provided the quotient is between .30 and 1.70. For this purpose use the months scales.



# IQ SLIDE RULE

## MENTAL AGE

36 Months  
3 Years

48  
4

60  
5

72  
6

84  
7

96  
8

108  
9

120  
10

132  
11

144  
12

156  
13

168  
14

180  
15

192  
16

204  
17

216  
18

110 120 130 140 150 160 170

30

40

50

60

70

80

90

IQ

Cut on these two lines

3 Years  
36 Months

48  
4

60  
5

72  
6

84  
7

96  
8

108  
9

120  
10

132  
11

144  
12

156  
13

168  
14

180  
15

192  
16

204  
17

216  
18

Read IQ here

CHRONOLOGICAL AGE

Read IQ here

Fold carefully on this line (line to be on outside of fold)

(See accompanying page for directions)

Cut on this line

Cut on this line





DATE DUE  
Unless this book is returned on ATE the last date stamp  
a fine will be charge  
ment of this LIBRARY

LB  
3063  
06

4854

**THE LIBRARY  
WEST VIRGINIA  
INSTITUTE OF TECHNOLOGY  
MONTGOMERY, WEST VIRGINIA 25136**



PRINTED IN U. S. A.

WVU Institute of Technology



1800103628



